ON SOME ASPECTS OF THE CLASSIFICATION PROBLEM

by

Hsien Elsa Lin

Technical Report No. 276

School of Statistics
University of Minnesota
Minneapolis, Minnesota 55455

October, 1976

This research was partially supported by a grant from the Mathematics Division, U.S. Army, Durham. (Grant No. DAAG 29-76-G-0038)

TABLE OF CONTENTS

INTRODUCTION

The problem of classification is to correctly associate one population π_0 with exactly one of several distinct populations $\pi_1, \ldots, \pi_m$. The problem may be to classify a single unit or more than one unit coming from π_0. The distribution functions which characterize the populations are not completely known. Sometimes the density functions are known except for some unknown parameters. In a broader problem the distribution functions are not given explicitly in simple parametric forms. In order to get more information on the distribution functions, data in the name of a "training sample" are collected in one of the following ways (depending on the situation):

(a) Separate samples from different populations.

(b) Sample from the population which is a mixture of $\pi_1, \ldots, \pi_m$.

When the density functions are known except for some parameters, a plug-in rule is obtained by replacing the parameters by the corresponding estimates (generally, maximum likelihood or some other consistent estimates are used) in the optimal rules according to some specified criteria in a given class. One may consider Bayes rules, minimax rules, admissible rules, etc. Asymptotic properties of most of these rules are not difficult to obtain, but asymptotic expansions of probabilities of misclassification (PMC) would be more useful. In Chapter one, we consider the problem of classifying one unit to one of three distinct multivariate normal distributions with a common but unknown covariance matrix. A plug-in rule is obtained by substituting the estimates of the parameters in the minimum distance (Mahalanobis distance) rule. Anderson (1973) obtained a similar result when $m=2$. Following T.W. Anderson (1973), we derive the asymptotic expansions of the PMC's and the estimated PMC's of this plug-in rule with an error of the order of the square of the number of observations. No such results are available in the literature for more than two populations.

When density functions are completely unknown, estimates of density functions are used to obtain a plug-in rule for a given rule which involves density functions. In 1951, Fix and Hodges proposed a classification rule for the two-population problem based on nonparametric estimates of the density functions. The K-nearest neighbor (K-NN) rule thus proposed by Fix and Hodges is described as follows: Let $\{X_{ij};\ j=1,\ldots,n_i\}$ be a random sample from the ith population. Consider a distance function d and order all the values $d(X_{ij}, Z)$ (Z is the observation to be classified.), $j=1,\ldots,n_i$; $i=1,\ldots,m$. The K-NN rule assigns Z to the population π_i, if $K_i/n_i = \max_j K_j/n_j$, where K_i is the number of pbservations from π_i in the K observations "nearest" to Z. They obtain the exact and asymptotic expressions for the PMC of the NN rule when $K = 1$. The 1-NN rule was also studied by Cover and Hart. Cover and Hart (1967) considered the mixed population case and proposed a K-NN rule which assigns Z to the population π_i, if $K_i = \max_j K_j$. In a recent paper, Goldstein (1972) has studied some asymptotic properties of the K_n-NN rules and obtained a consistent upper bound for its PMC. In Chapter two, we propose some rules which use the basic ideas of the NN rules, but are expressed in terms of their ranks, when the observations are available only in their relative orders or ranks and the usual NN rules can't be applied. However, it may be noted that the density functions can't be estimated using the ranked observations only. The asymptotic PMC's of these rules are derived and when sampling from a mixed population is considered, asymptotic risks are obtained as well. The asymptotic risk of the modified 1-NN rule is the same as the respective asymptotic risk of the 1-NN rule. The asymptotic risk of the modified K_n-NN rule turns out to be exactly the Bayes risk.

Another class of rules are suggested based on U-statistics. Das Gupta (1964) proposed a rule based on Wilcoxon statistics. He showed that such a rule is consistent. Hudimoto (1964) also used Wilcoxon statistic when $\int F_1 dF_2 - \frac{1}{2} > 0$ and derived some bounds for the probability of error. Chanda and Lee (1975) modified Hudimoto's rule to the situation when either $\int F_1 dF_2 - \frac{1}{2} > 0$ or $\int F_1 dF_2 - \frac{1}{2} < 0$. We shall use Hudimoto's idea and suggest a two-sided classification rule based on the Lehmann statistic (Lehmann, 1951). Asymptotic results are of theoretical interest; however good studies on the rate of convergence will be useful. Following Grams and Serfling (1973) in their study of convergence rate for U-statistics, we obtain the asymptotic PMC's of these rules, together with the rate of convergence when the sizes of the training samples approach infinity. The strong consistency of the rules are also pointed out.

Finally, we consider sequential rules in order to attain prescribed probabilities of error. Hoeffding and Wolfowitz (1958) studied the problem of distinguishability of sets of distributions. Later the notion of distinguishability was used by Das Gupta and Kinderman (1974) in the set-up for the classification problems. Hoeffding and Wolfowitz (1958) introduced the minimum distance test procedure and studied the properties of this test using the available probability bounds on sample distribution function. In Chapter four, we shall introduce the minimum-U sequential rules and prove some properties of these rules by using the available probability inequality for U-statistics. Srivastava (1973) considered sequential rules for classification into one of two distinct multivariate normal distribution with means μ_1, μ_2 and a common covariance matrix Σ in the following two cases: (i) $\mu_1-\mu_2 = \delta$ is known but Σ is unknown. (ii) Both δ

and Σ are unknown. For the case (i) Srivastava proposed a sequential rule based on observations from Π_0 and Π_1, given α he showed that the PMC's of this rule tend to values less than α as $\delta' \Sigma^{-1}\delta \to 0$. However Srivastava's proof is incomplete and suffers from a technical error. We shall present a more rigorous analysis of his rule. Srivastava also proved that for his rule in case (ii) the error can be controlled arbitrarily as $\delta' \Sigma^{-1}\delta \to 0$. But his proof is entirely wrong and we shall indicate his error.

CHAPTER I

CLASSIFICATION INTO ONE OF THREE MULTIVARIATE NORMAL DISTRIBUTIONS

1.0 Introduction

A random observation X is drawn from $N_p(\mu, \Sigma)$. The problem is to classify this distribution into one of $N_p(\mu_1, \Sigma)$, $N_p(\mu_2, \Sigma)$, and $N_p(\mu_3, \Sigma)$. It is assumed that μ_1, μ_2, and μ_3 are distinct and Σ is nonsingular (see T. W. Anderson (1958), Chapter 6). When μ_1, μ_2, μ_3 and Σ are known, and the costs of misclassification are equal then under the assumption that drawing a new observation from each population is equally likely the optimal classification rule (minimizing the expected loss from cost of misclassification) δ decides $\mu = \mu_i$, iff $(X-\mu_i)'\Sigma^{-1}(X-\mu_i) = \min_{j=1,2,3} (X-\mu_j)'\Sigma^{-1}(X-\mu_j)$, which may be written as

$$(1.1) \quad U_{ij} = (X-\tfrac{1}{2}(\mu_i+\mu_j))'\Sigma^{-1}(\mu_i-\mu_j) > 0 \quad \text{and}$$

$$U_{ik} = (X-\tfrac{1}{2}(\mu_i+\mu_k))'\Sigma^{-1}(\mu_i-\mu_k) > 0$$

where $i, j, k = 1, 2, 3$; $i \neq j$, $j \neq k$, $k \neq i$.

To compute the PMC's of this rule, let us assume that $\mu = \mu_1$. Then

$$(1.2)\quad \pi_{21} = \Pr(\delta \text{ decides } \mu = \mu_2 | \mu = \mu_1)$$
$$= \Pr(U_{21} > 0,\ U_{23} > 0 | \mu = \mu_1)$$

Let

$$(1.3)\quad \alpha_{ij} = (\mu_1-\mu_i)'\Sigma^{-1}(\mu_1-\mu_j), \qquad i,j = 2,3.$$

Then

$$(1.4)\quad \pi_{21} = \Psi(\alpha, \beta; \rho),$$

where

$$(1.5)\quad \alpha = \tfrac{1}{2}\alpha_{22}^{\frac{1}{2}},\ \beta = \tfrac{1}{2}(\alpha_{22}-\alpha_{33})/(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{\frac{1}{2}},$$

$$(1.6)\quad \rho = (\alpha_{22}-\alpha_{23})/[\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})]^{\frac{1}{2}}$$

and

$$(1.7)\quad \Psi(\alpha,\beta;\rho) = \int_{\alpha}^{\infty}\int_{\beta}^{\infty}\varphi_2(u,v;\rho)\,du\,dv,$$

$\varphi_2(.,.;\rho)$ being the pdf of the bivariate normal distribution with zero means, unit variances and correlation coefficient ρ. The PMC π_{31} can be obtained by interchanging the subscripts 2 and 3 in the formula for π_{21}.

But in most applications the parameters are not known and a training sample from each population is available: $(X_{i1}, X_{i2}, \ldots, X_{in_i})$ is drawn from $N_p(\mu_i, \Sigma)$, $i = 1,2,3$. Estimates (based on training samples) of the parameters are substituted in (1.1). To get a rule called a plug-in version of δ we estimate μ_i by

$$(1.8)\quad \bar{X}_i = \frac{1}{n_i}\sum_{\alpha=1}^{n_i} X_{i\alpha}, \qquad i=1,2,3$$

and Σ by S, where

$$(1.9)\quad (n_1+n_2+n_3-3)S = \sum_{i=1}^{3}\sum_{\alpha=1}^{n_i}(X_{i\alpha}-\bar{X}_i)(X_{i\alpha}-\bar{X}_i)'.$$

Then the plug-in minimum distance rule $\hat{\delta}$ decides $\mu = \mu_i$, iff

$$(1.10)\quad (X-\tfrac{1}{2}(\bar{X}_i+\bar{X}_j))'S^{-1}(\bar{X}_i-\bar{X}_j) > 0 \quad\text{and}$$

$$(X-\tfrac{1}{2}(\bar{X}_i+\bar{X}_k))'S^{-1}(\bar{X}_i-\bar{X}_k) > 0.$$

In this chapter, we obtain asymptotic expansions of the PMC's and the estimated PMC's of the plug-in rule with an error of the order of the square of the number of observations. No such results are available in the literature for more than two populations. Anderson (1973) obtained similar results for the two-population problem.

1.1 The asymptotic expansions of PMC's.

The PMC's of the plug-in rule $\hat{\delta}$ will be derived now under the assumption $\mu = \mu_1$.

$$(1.11)\quad P_{21} = \Pr(\hat{\delta} \text{ decides } \mu = \mu_2 | \mu = \mu_1)$$

$$= \Pr(\hat{U}_{21} > 0,\ \hat{U}_{23} > 0 | \mu = \mu_1),$$

where $\hat{U}$'s are obtained from the corresponding U's after replacing μ_1, μ_2, μ_3, Σ by $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$, S, respectively. Conditioning on $\bar{X}_i$'s and S we get

$$(1.12)\ P_{21}(\bar{X}_1, \bar{X}_2, \bar{X}_3, S)$$
$$\equiv \Psi(-(\mu_1-\tfrac{1}{2}(\bar{X}_2+\bar{X}_1))'S^{-1}(\bar{X}_2-\bar{X}_1)/[(\bar{X}_2-\bar{X}_1)'S^{-1}\Sigma S^{-1}(\bar{X}_2-\bar{X}_1)]^{\frac{1}{2}},$$
$$-(\mu_1-\tfrac{1}{2}(\bar{X}_2+\bar{X}_3))'S^{-1}(\bar{X}_2-\bar{X}_3)/[(\bar{X}_2-\bar{X}_3)'S^{-1}\Sigma S^{-1}(\bar{X}_2-\bar{X}_3)]^{\frac{1}{2}};$$
$$(\bar{X}_2-\bar{X}_1)'S^{-1}\Sigma S^{-1}(\bar{X}_2-\bar{X}_3)/[(\bar{X}_2-\bar{X}_1)'S^{-1}\Sigma S^{-1}(\bar{X}_2-\bar{X}_1)$$
$$(\bar{X}_2-\bar{X}_3)'S^{-1}\Sigma S^{-1}(\bar{X}_2-\bar{X}_3)]^{\frac{1}{2}}).$$

For simplicity, we shall assume $n = n_1 = n_2 = n_3$ and from now on write

$$(1.13)\ m = n_1+n_2+n_3-3 = 3n-3.$$

The distribution of $(\hat{U}_{21}, \hat{U}_{23})$ is invariant with respect to the transformations $X^* = AX+b$, $X^*_{ij} = AX_{ij}+b$, $j = 1,\dots,n$; $i = 1,2,3$, where A is nonsingular. Without loss of generality, we shall replace μ_1, μ_2, μ_3 and Σ by 0, η_2, η_3 and I, respectively, where

$$(1.14)\ \eta_2 = \Sigma^{-\frac{1}{2}}(\mu_1-\mu_2), \qquad \eta_3 = \Sigma^{-\frac{1}{2}}(\mu_1-\mu_3).$$

Then

$$(1.15)\ \alpha_{22} = \eta_2'\eta_2, \quad \alpha_{33} = \eta_3'\eta_3, \quad \alpha_{23} = \eta_2'\eta_3.$$

Define Y_1, Y_2, Y_3 and V by

$$(1.16)\quad \bar{X}_1 = Y_1/m^{\frac{1}{2}},\quad \bar{X}_2 = -\eta_2+Y_2/m^{\frac{1}{2}},\quad \bar{X}_3 = -\eta_3+Y_3/m^{\frac{1}{2}},$$
$$S = I+V/m^{\frac{1}{2}}.$$

The statistics X, $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$, and S are independently distributed as $N_p(\mu_1, \Sigma)$, $N_p(\mu_2, \Sigma/n)$, $N_p(\mu_3, \Sigma/n)$, and $W(\Sigma, m)$, respectively. Combining these and the transformations mentioned above, we can assume that X, Y_1, Y_2, Y_3 and V are mutually independent and $X \sim N_p(0, I)$; $Y_j \sim N_p(0\ mI/n)$, $i = 1,2,3$; and $\mathcal{E}V = 0$. Then, in terms of Y's and V, (1.12) is

$$(1.17)\quad P_{21}(Y_1,Y_2,Y_3,V) \equiv \Psi(a_m, b_m;r_m),$$

where

$$(1.18)\quad a_m = \tfrac{1}{2}[-\eta_2+(Y_2+Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]/$$
$$\{[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-2}[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]\}^{\frac{1}{2}},$$

$$(1.19)\quad b_m = \tfrac{1}{2}[-(\eta_2+\eta_3)+(Y_2+Y_3)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}$$
$$[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]/\{[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]'$$
$$(I+V/m^{\frac{1}{2}})^{-2}[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]\}^{\frac{1}{2}},$$

$$(1.20)\quad r_m = [-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-2}[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]/$$
$$\{[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-2}[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]$$
$$[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-2}[-(\eta_2-\eta_3)$$
$$+(Y_2-Y_3)/m^{\frac{1}{2}}]\}^{\frac{1}{2}}.$$

Note that

$$(1.21)\quad (I+V/m^{\frac{1}{2}})^{-1} = I-V/m^{\frac{1}{2}}+V^2/m^{3/2}-V^3/m^{3/2}+V^4/m^2+V^5(I+V/m^{\frac{1}{2}})^{-1}/m^{5/2}$$

$$(1.22)\quad (I+V/m^{\frac{1}{2}})^{-2} = I-2V/m^{\frac{1}{2}}+3V^2/m-4V^3/m^{3/2}+5V^4/m^{5/2}$$
$$-(6V^5+5V^6/m^{\frac{1}{2}})(I+V/m^{\frac{1}{2}})^{-2}/m^{5/2}.$$

Let J_m be the subset of the sample space defined as

$$(1.23)\quad J_m = \{|Y_{kj}| < g(\log m)^{\frac{1}{2}},\ |V_{ij}| < 2\log m;\ k = 1,2,3,$$
$$i,j = 1, \ldots, p, \quad \text{is a constant greater than } 4\},$$

where $Y_k = (Y_{k1}, \ldots, Y_{kp})'$. Lemma of Anderson (1973) yields the following:

<u>Lemma 1.1</u> $\Pr(J_m) = 1-o(m^{-2})$.

Consequently, since $0 \leq \Psi(G_m,b_m;r_m) \leq 1$ we have

$$(1.24)\quad P_{21} = \mathcal{E}\Psi(G_m,b_m;r_m) = \mathcal{E}\Psi(G_m,b_m;r_m)\chi(J_m)$$
$$+ \mathcal{E}\Psi(G_m,b_m;r_m)\chi(J_m^{\,c})$$
$$= \mathcal{E}\Psi(G_m,b_m;r_n)\chi(J_m)+O(m^{-2}),$$

where $\chi(A)$ stands for the indicator set function of a set A.

Define

$$(1.25)\quad a = \alpha_{22},\ b = \alpha_{33},\ d = \alpha_{23},$$

$$e = \alpha_{22}-\alpha_{33},\ f = \alpha_{22}-\alpha_{23},\ q = \alpha_{33}-\alpha_{23},$$

$$\sigma^2 = \alpha_{22}+\alpha_{33}-2\alpha_{23},$$

$$\xi = \alpha_{22}\alpha_{23}+\alpha_{33}\alpha_{23}-2\alpha_{22}\alpha_{33}$$

$$\tau^2 = \alpha_{22}\alpha_{33}-\alpha_{23}^2.$$

Using the identities (1.21) and (1.22), over J_m for sufficiently large m (as Taylor series expansions), we have the following:

The numerator of a_m in (1.18) is $\frac{1}{2}$ times

$$(1.26)\quad a+\frac{1}{m^{\frac{1}{2}}}(-\eta_2'V\eta_2-2\eta_2'Y_2)$$
$$+\frac{1}{m}[\eta_2'V^2\eta_2+2\eta_2'VY_2+(Y_2+Y_1)'(Y_2-Y_1)]+\gamma_{1m}(Y_1,Y_2,V),$$

where $\gamma_{1m}(Y_1,Y_2,V)$ is a remainder term consisting of $m^{-3/2}$ times a homogeneous polynomial (not depending on m) of degree 3 in the elements of Y_1,Y_2,Y_3, and V plus m^{-2} times in a homogeneous polynomial of degree 4 plus a remainder term which is $O(m^{-5/2})$ for fixed Y_1,Y_2,Y_3, and V (and $O(\log m)^5/m^{5/2}$) for Y_1,Y_2,Y_3, and V in J_m).

The denominator of a_m in (1.18) is

$$(1.27)\quad \{\eta_2'\eta_2 + \frac{1}{m^{\frac{1}{2}}}[-2\eta_2' V\eta_2 - 2\eta_2'(Y_2-Y_1)]$$

$$+ \frac{1}{m}[3\eta_2' V^2\eta_2 + 4\eta_2'(Y_2-Y_1) + (Y_2-Y_1)'(Y_2-Y_1) + \gamma_{2m}(Y_1,Y_2,V)]\}^{-\frac{1}{2}}$$

$$= \frac{1}{G^{\frac{1}{2}}} + \frac{1}{m^{\frac{1}{2}}}[\eta_2'(Y_2-Y_1) + \eta_2' V\eta_2]/G^{3/2}$$

$$+ \frac{1}{m}\{[3\eta_2' V^2\eta_2 + 4\eta_2' V(Y_2-Y_1) + (Y_2-Y_1)'(Y_2-Y_1)]/2G^{3/2}$$

$$+3[\eta_2'(Y_2-Y_1) + \eta_2' V\eta_2]^2/2G^{5/2}\} + \gamma_{3m}(Y_1,Y_2,V),$$

where $\gamma_{2m}(Y_1,Y_2,V)$, $\gamma_{3m}(Y_1,Y_2,V)$ have the same properties as $\gamma_{1m}(Y_1,Y_2,V)$. The notation $\gamma_{jm}(Y_1,Y_2,Y_3,V)$ will be used frequently, which will have the same properties as those of γ_{1m}, unless mentioned otherwise.

Combining (1.26) and (1.27), we get

$$(1.28)\quad G_m = \alpha + C/m^{\frac{1}{2}} + D/m + \gamma_{4m}(Y_1,Y_2,V),$$

where

$$(1.29)\quad C = -\eta_2'(Y_2+Y_1)/2G^{\frac{1}{2}},$$

$$D = [-\eta_2' V^2\eta_2 + 4\eta_2' VY_1 + (Y_2+3Y_1)'(Y_2-Y_1)]/4G^{\frac{1}{2}}$$

$$+[(\eta_2' V\eta_2)^2 - 4\eta_2' V\eta_2\eta_2' Y_1 - \eta_2'(Y_2+3Y_1)(Y_2-Y_1)'\eta_2]/4G^{3/2}.$$

The numerator of b_m in (1.19) is $\frac{1}{2}$ times

$$(1.30)\quad e + \frac{1}{m^{\frac{1}{2}}}[-(\eta_2+\eta_3)' V(\eta_2-\eta_3) - 2(\eta_2' Y_2 - \eta_3' Y_3)]$$

$$+ \frac{1}{m}[(\eta_2+\eta_3)' V^2(\eta_2-\eta_3) + 2(\eta_2' VY_2 - \eta_3' VY_3) + (Y_2+Y_3)'(Y_2-Y_3)]$$

$$+ \gamma_{5m}(Y_2,Y_3,V).$$

The denominator of b_m in (1.19) is

$$(1.31)\quad \frac{1}{\sigma}+\frac{1}{m^{\frac{1}{2}}}[(\eta_2-\eta_3)'(Y_2-Y_3)+(\eta_2-\eta_3)'V(\eta_2-\eta_3)]/\sigma^3$$
$$+\frac{1}{m}\{-[3(\eta_2-\eta_3)'V^2(\eta_2-\eta_3)+(Y_2-Y_3)'(Y_2-Y_3)$$
$$+4(\eta_2-\eta_3)'V(Y_2-Y_3)]/2\sigma^3$$
$$+3[(\eta_2-\eta_3)'(Y_2-Y_3)+(\eta_2-\eta_3)'V(\eta_2-\eta_3)]^2/2\sigma^5\}$$
$$+\gamma_{6m}(Y_2,Y_3,V).$$

Combining (1.30) and (1.31), we get

$$(1.32)\quad b_m=\beta+F/m^{\frac{1}{2}}+G/m+\gamma_{7m}(Y_2,Y_3,V),$$

where

$$(1.33)\quad F=-[(\eta_2+\eta_3)'V(\eta_2-\eta_3)+2(\eta_2'Y_2-\eta_3'Y_3)]/2\sigma$$
$$+e[(\eta_2-\eta_3)/(Y_2-Y_3)+(\eta_2-\eta_3)'V(\eta_2-\eta_3)]/2\sigma^3$$
$$G=-e[3(\eta_2-\eta_3)'V^2(\eta_2-\eta_3)+(Y_2-Y_3)'(Y_2-Y_3)$$
$$+4(\eta_2-\eta_3)'V(Y_2-Y_3)]/4\sigma^3+3e[(\eta_2-\eta_3)'(Y_2-Y_3)$$
$$+(\eta_2-\eta_3)'V(\eta_2-\eta_3)]^2/4\sigma^5+[(\eta_2+\eta_3)'V^2(\eta_2-\eta_3)$$
$$+2(\eta_2'VY_2-\eta_3'VY_3)+(Y_2+Y_3)'(Y_2-Y_3)]/2\sigma$$
$$-[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+(\eta_2-\eta_3)'(Y_2-Y_3)][(\eta_2+\eta_3)'V(\eta_2-\eta_3)$$
$$+2(\eta_2Y_2-\eta_3Y_3)]/2\sigma^3$$

The numerator of r_m in (1.20) is

$$(1.34)\quad f + \frac{1}{m^{\frac{1}{2}}}[-\eta_2' V(\eta_2-\eta_3)-\eta_2'(Y_2-Y_3)-(\eta_2'-\eta_3)'(Y_2-Y_1)]$$

$$+ \frac{1}{m}[3\eta_2' V^2(\eta_2-\eta_3)+2\eta_2' V(Y_2-Y_3)+2(\eta_2-\eta_3)' V(Y_2-Y_1)$$

$$+(Y_2-Y_1)'(Y_2-Y_3)]+ \gamma_{8m}(Y_1,Y_2,Y_3,V).$$

Combining (1.34), (1.27), and (1.31), we get

$$(1.35)\quad r_m = \rho+H/m^{\frac{1}{2}}+K/m+\gamma_{9m}(Y_1,Y_2,Y_3,V),$$

where

$$(1.36)\quad H = -[2\eta_2' V(\eta_2-\eta_3)-\eta_2'(Y_2-Y_3)-(\eta_2-\eta_3)'(Y_2-Y_1)]/G^{\frac{1}{2}}\sigma$$

$$+f[\eta_2'(Y_2-Y_1)+\eta_2' V\eta_2]/G^{3/2}\sigma$$

$$+f[(\eta_2-\eta_3)'(Y_2-Y_3)+(\eta_2-\eta_3)' V(\eta_2-\eta_3)]/G^{\frac{1}{2}}\sigma^3$$

$$K = -f[3(\eta_2-\eta_3)' V^2(\eta_2-\eta_3)+(Y_2-Y_3)'(Y_2-Y_3)$$

$$+4(\eta_2-\eta_3)' V(Y_2-Y_3)]/2G^{\frac{1}{2}}\sigma^3$$

$$+3f[(\eta_2-\eta_3)'(Y_2-Y_3)+(\eta_2-\eta_3)' V(\eta_2-\eta_3)]^2/2G^{\frac{1}{2}}\sigma^5$$

$$+[3\eta_2' V^2(\eta_2-\eta_3)+2\eta_2' V(Y_2-Y_3)+2(\eta_2-\eta_3)' V(Y_2-Y_1)$$

$$+(Y_2-Y_1)'(Y_2-Y_3)]/G^{\frac{1}{2}}\sigma$$

$$-f[3\eta_2' V^2\eta_2+(Y_2-Y_1)'(Y_2-Y_1)+4\eta_2' V(Y_2-Y_1)]/2G^{3/2}\sigma$$

$$+3f[\eta_2'(Y_2-Y_1)+\eta_2' V\eta_2]^2/2G^{5/2}\sigma$$

$$+f[\eta_2'(Y_2-Y_1)+\eta_2' V\eta_2][(\eta_2-\eta_3)'(Y_2-Y_3)$$

$$+(\eta_2-\eta_3)' V(\eta_2-\eta_3)]/G^{3/2}\sigma^3$$

$$+[2\eta_2'V(\eta_2-\eta_3)+\eta_2'(Y_2-Y_3)+(\eta_2-\eta_3)'(Y_2-Y_1)] \times$$

$$[(\eta_2-\eta_3)'(Y_2-Y_3)+(\eta_2-\eta_3)'V(\eta_2-\eta_3)]/G^{\frac{1}{2}}\sigma^3$$

$$-[2\eta_2'V(\eta_2-\eta_3)+\eta_2'(Y_2-Y_3)+(\eta_2-\eta_3)'(Y_2-Y_1)] \times$$

$$[\eta_2'(Y_2-Y_1)+\eta_2'V\eta_2]/G^{3/2}\sigma.$$

We assume that $|\rho| \neq 1$. This means that μ_1, μ_2, μ_3 are not linearly related.

Then a Taylor series expansion of $\Psi(G_m, b_m; r_m)$ over J_m for sufficiently large m (see appendix for detailed derivation) gives,

(1.37) $\Psi(G_m, b_m; r_m)$

$$= \Psi(\alpha,\beta;\rho)+\varphi_1(\alpha)\Phi_1(G^{\frac{1}{2}}q/2\tau)[-C/m^{\frac{1}{2}}+G^{\frac{1}{2}}C^2/4m-D/m]$$

$$+\varphi_1(\beta)\Phi_1(\xi/2\sigma\tau)[eF^2/4m\sigma-F/m^{\frac{1}{2}}-G/m]$$

$$+\varphi_1(\alpha)\varphi_1(-G^{\frac{1}{2}}q/2\tau)(G^{\frac{1}{2}}\sigma/\tau)[-fC^2/2mG^{\frac{1}{2}}\sigma+CF/2m+G^{\frac{1}{2}}\xi CH/4m\tau^2]$$

$$+\varphi_1(\beta)\varphi_1(-\xi/2\sigma\tau)(G^{\frac{1}{2}}\sigma/\tau)[-fF^2/2mG^{\frac{1}{2}}\sigma+CF/2m+G\sigma qFH/4m\tau^2]$$

$$+\varphi_2(\alpha,\beta;\rho)\{H/m^{\frac{1}{2}}+K/m+(G\sigma/2\tau^2)[f/G^{\frac{1}{2}}+G^{\frac{1}{2}}q\xi/4\tau^2]H^2/m$$

$$+(G^{\frac{1}{2}}\xi/4\tau^2)CH/m+(Gq\sigma/4\tau^2)FH/m\}+\gamma_{10}(Y_1,Y_2,Y_3,V)/m^{3/2}$$

$$+\gamma_{11}(Y_1,Y_2,Y_3,V)+\gamma_{12m}(Y_1,Y_2,Y_3,V),$$

where $\Phi_1(x) = \int_{-\infty}^{x}\varphi_1(y)dy$; and

$\varphi_1(\cdot)$ being the pdf of the standard normal distribution; and $\gamma_{10}(Y_1,Y_2,Y_3,V)$ is a homogeneous polynomial (not depending on

m) of degree 3 in the elements of Y_1, Y_2, Y_3, and V;
$\gamma_{11}(Y_1, Y_2, Y_3, V)$ is a polynomial of degree 4, and
$\gamma_{12m}(Y_1, Y_2, Y_3, V)$ is a remainder term, which is $O(m^{-5/2})$ for fixed Y_1, Y_2, Y_3, and V in J_m.

Since J_m is by definition symmetric in Y_1, Y_2, and Y_3, C has the expectation zero over J_m. Let h be a function of Y_1, Y_2, Y_3, and V having finite second moment. Then

$$(1.38)\quad \begin{aligned} &|\mathcal{E}h - \mathcal{E}h\chi(J_m)| \\ &= |\mathcal{E}h\chi(J_m^{\,c})| \\ &\leq |\mathcal{E}h^2|^{\frac{1}{2}} |\mathcal{E}\chi(J_m^{\,c})|^{\frac{1}{2}} \\ &= O(m^{-1}). \end{aligned}$$

Consequently, the differences between $\mathcal{E}D/m$, $\mathcal{E}G/m$, $\mathcal{E}K/m$, $\mathcal{E}C^2/m$, $\mathcal{E}F^2/m$, $\mathcal{E}H^2/m$, $\mathcal{E}CF/m$, $\mathcal{E}CH/m$, $\mathcal{E}FH/m$, $\mathcal{E}\gamma_{10}(Y_1, Y_2, Y_3, V)/m^{3/2}$ and the corresponding expectations over J_m are $o(m^{-2})$. Moreover, for any positive integer t,

$$(1.39)\quad \mathcal{E}V_{ij}^{\,2t-1} = m^{t-\frac{1}{2}} O(m^{-t}) = O(m^{-\frac{1}{2}})$$

Hence

$$(1.40)\quad \begin{aligned} &|\mathcal{E}F/m^{\frac{1}{2}} - \mathcal{E}F\chi(J_m)/m^{\frac{1}{2}}| \\ &= |\mathcal{E}F\chi(J_m^{\,c})/m^{\frac{1}{2}}| \\ &\leq (1/m^{\frac{1}{2}}) |\mathcal{E}F^3|^{1/3} |\mathcal{E}\chi(J_m^{\,c})|^{2/3} \\ &= (1/m^{\frac{1}{2}}) |O(m^{-\frac{1}{2}}))^{1/3} (o(m^{-2}))^{2/3} \\ &= o(m^{-2}). \end{aligned}$$

Similarly,

$$|\mathcal{E}H/m^{\frac{1}{2}}-\mathcal{E}H\chi(J_m)/m^{\frac{1}{2}}| \leq o(m^{-2}).$$

Note also that $\mathcal{E}F = 0$, $\mathcal{E}H = 0$. Since the fourth-order absolute moments of Y_1, Y_2, Y_3, and V exist and are bounded, so is $\mathcal{E}\gamma_{11}(Y_1,Y_2,Y_3,V)\chi(J_m)$. Hence $(1/m^2)\mathcal{E}\gamma_{11}(Y_1,Y_2,Y_3,V)\chi(J_m)=O(m^{-2})$. Finally, in J_m each element of Y_1, Y_2, Y_3, and V divided by $m^{\frac{1}{2}}$ is less than a constant times $\log m/m^{\frac{1}{2}}$, therefore

$$(1.41)\quad \mathcal{E}|\gamma_{12m}(Y_1,Y_2,Y_3,V)\chi(J_m)| = O(m^{-5/2}\log^5 m) = o(m^{-2}).$$

Thus

$$(1.42)\quad \mathcal{E}\Psi(G_m,b_m;r_m)\chi(J_m) = \Psi(\alpha,\beta;\rho)+Q/m+\mathcal{E}\gamma_{10}(Y_1,Y_2,Y_3,V)/m^{3/2} +O(m^{-2}),$$

where

$$\begin{aligned}(1.43)\quad Q = {} & \varphi_1(\alpha)\Phi_1(G^{\frac{1}{2}}q/2\tau)\mathcal{E}[G^{\frac{1}{2}}C^2/4-D]+\varphi_1(\beta)\Phi_1(\xi/2\sigma\tau)\mathcal{E}[eF^2/4\sigma-G] \\ & +\varphi_1(\alpha)\varphi_1(-G^{\frac{1}{2}}q/2\tau)(G^{\frac{1}{2}}\sigma/\tau)\mathcal{E}[-fC^2/2G^{\frac{1}{2}}\sigma+CF/2-G^{\frac{1}{2}}\xi CH/4\tau^2] \\ & +\varphi_1(\beta)\varphi_1(-\xi/2\sigma\tau)(G^{\frac{1}{2}}\sigma/\tau)\mathcal{E}[-fF^2/2G^{\frac{1}{2}}\sigma+CF/2+G\sigma qFH/4\tau^2] \\ & +\varphi_2(\alpha,\beta;\rho)\mathcal{E}\{K+G\sigma/2\tau^2[f/G^{\frac{1}{2}}-G^{\frac{1}{2}}q\xi/4\tau^2]H^2+G^{\frac{1}{2}}\xi CH/4\tau^2 \\ & +G\sigma qFH/4\tau^2\}\end{aligned}$$

Since the third moments of Y_1, Y_2, Y_3, and V are either zero or $O(m^{-\frac{1}{2}})$, combining (1.24) and (1.42), we have

$$(1.44)\quad P_{21} = \Psi(\alpha,\beta;\rho)+Q/m+O(m^{-2}).$$

Finally we have (see Appendix for details),

$$\begin{aligned}(1.45)\quad P_{21} = {} & \Psi(\alpha,\beta;\rho) \\ & +\frac{1}{m}\Big\{\varphi_1(\alpha)\Phi_1(\mathfrak{a}^{\frac{1}{2}}q/2\tau)[(2p+1)\mathfrak{a}^{\frac{1}{2}}/8+3(p-1)/2\mathfrak{a}^{\frac{1}{2}}] \\ & +\varphi_1(\beta)\Phi_1(\xi/2\sigma\tau)(e/4\sigma)[(p-1)/4+(\tau^2+3(\mathfrak{a}+b)+6(p-1))/\sigma^2 \\ & -3(\mathfrak{a}-b)^2/2\sigma^4]-(3/8)\varphi_1(\alpha)\varphi_1(\beta)(f-d)/\tau \\ & +\varphi_1(\beta)\varphi_1(-\xi/2\sigma\tau)(1/\tau\sigma^2)[-\tfrac{1}{2}f(\tau^2-3e^2/2\sigma^2)+(1/8)(-9\mathfrak{a}^2 \\ & +8\mathfrak{a}q+5b^2+4d^2)]+\varphi_2(\alpha,\beta;\rho)\{(1/\mathfrak{a}^{\frac{1}{2}}\sigma)[3(p-1)-3(p-1)q/\mathfrak{a} \\ & -3(p-1)f/\sigma^2-2f+f^2(3+2f)/\mathfrak{a}\sigma^2]+\xi/8\mathfrak{a}^{\frac{1}{2}}\sigma^3 \\ & +q(11\mathfrak{a}+5b-4d)/8\mathfrak{a}^{\frac{1}{2}}\sigma^3+(1/\sigma\tau^2)[f/\mathfrak{a}^{\frac{1}{2}}+\mathfrak{a}^{\frac{1}{2}}q\xi/4\tau^2][-4f^2 \\ & -3f-3f^2/\mathfrak{a}-3f^2/\sigma^2+3\sigma^2+3\mathfrak{a}+2\mathfrak{a}\sigma^2 \\ & -2f^3(3+2f)/\mathfrak{a}\sigma^2]\}\Big\}+O(m^{-2}).\end{aligned}$$

The asymptotic expansion of the PMC P_{31} can be obtained by interchanging the subscripts 2 and 3 in (1.45).

1.2 The asymptotic expansions of estimated PMC's.

We estimate P_{21} by considering X distributed as $N_p(\bar{X}_1,S)$. Then, in terms of Y's and V, $\hat{P}_{21}(Y_1,Y_2,Y_3,V)$ is

$$(1.46)\quad \hat{P}_{21}(Y_1,Y_2,Y_3,V) = \Psi(\hat{\mathfrak{a}}_m,\hat{b}_m;\hat{r}_m),$$

where

$$(1.47)\quad \hat{\sigma}_m = \tfrac{1}{2}\{[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]\}^{\frac{1}{2}},$$

$$(1.48)\quad \hat{b}_m = \tfrac{1}{2}[-(\eta_2+\eta_3)+(Y_2+Y_3-2Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}[-(\eta_2+\eta_3)$$
$$+(Y_2+Y_3-2Y_1)/m^{\frac{1}{2}}]/\{[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}\times$$
$$[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]\}^{\frac{1}{2}},$$

$$(1.49)\quad \hat{r}_m = [-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]/$$
$$\{[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}[-\eta_2+(Y_2-Y_1)/m^{\frac{1}{2}}][-(\eta_2-\eta_3)$$
$$+(Y_2-Y_3)/m^{\frac{1}{2}}]'(I+V/m^{\frac{1}{2}})^{-1}[-(\eta_2-\eta_3)+(Y_2-Y_3)/m^{\frac{1}{2}}]\}^{\frac{1}{2}}.$$

As before, over J_m for sufficiently large m, Taylor series expansions give:

$$(1.50)\quad \hat{\sigma}_m = \alpha + C^*/m^{\frac{1}{2}}+D^*/m + \gamma_{13m}(Y_1,Y_2,V),$$

where

$$(1.51)\quad C^* = -[\eta_2'V\eta_2+2\eta_2'(Y_2-Y_1)]/4\alpha^{\frac{1}{2}}.$$
$$D^* = [\eta_2'V^2\eta_2+2\eta_2'V(Y_2-Y_1)+(Y_2-Y_1)'(Y_2-Y_1)]/4\alpha^{\frac{1}{2}}$$
$$-[\eta_2'V\eta_2+2\eta_2'(Y_2-Y_1)]^2/16\alpha^{3/2}.$$

$$(1.52)\quad \hat{b}_m = \beta + F^*/m^{\frac{1}{2}} +G^*/m + \gamma_{14m}(Y_1,Y_2,Y_3,V),$$

where

$$(1.53)\quad F^* = [-\eta_2'V\eta_2+\eta_3'V\eta_3-2\eta'_2(Y_2-Y_1)+2\eta_3'(Y_3-Y_1)]/2\sigma$$
$$+e[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+2(\eta_2-\eta_3)'(Y_2-Y_3)]/4\sigma^3,$$

$$G^* = -e[(\eta_2-\eta_3)'V^2(\eta_2-\eta_3)+2(\eta_2-\eta_3)'V(Y_2-Y_3)+(Y_2-Y_3)'(Y_2-Y_3)]/4\sigma^3$$
$$+3e[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+2(\eta_2-\eta_3)'(Y_2-Y_3)]^2/16\sigma^5$$
$$+[-\eta_2'V\eta_2+\eta_3'V\eta_3-2\eta_2'(Y_2-Y_1)+2\eta_3'(Y_3-Y_1)] \times [(\eta_2-\eta_3)'V(\eta_2-\eta_3)+2(\eta_2-\eta_3)'(Y_2-Y_3)]/2\sigma^3.$$

$$(1.54)\quad \hat{r}_m = \rho + H^*/m^{\frac{1}{2}} + K^*/m + \gamma_{15m}(Y_1,Y_2,Y_3,V),$$

where

$$(1.55)\quad H^* = f[\eta_2'V\eta_2+2\eta_2'(Y_2-Y_1)]/2G^{3/2}\sigma+[-\eta_2'V\eta_2+\eta_2'V\eta_3$$
$$-\eta_2'(Y_2-Y_1)+\eta_2'(Y_3-Y_1)+\eta_3'(Y_2-Y_1)]/G^{\frac{1}{2}}\sigma$$
$$+f[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+2(\eta_2-\eta_3)'(Y_2-Y_3)]/2G^{\frac{1}{2}}\sigma^3$$

$$K = [\eta_2'V^2(\eta_2-\eta_3)+2\eta_2'V(Y_2-Y_1)-\eta_2'V(Y_3-Y_1)-\eta_3'(Y_2-Y_1)$$
$$+(Y_2-Y_1)'(Y_2-Y_1)-(Y_3-Y_1)'(Y_3-Y_1)]/G^{\frac{1}{2}}\sigma$$
$$-f[(\eta_2-\eta_3)'V^2(\eta_2-\eta_3)+2(\eta_2-\eta_3)'V(Y_2-Y_3)+(Y_2-Y_3)'(Y_2-Y_3)]/2G^{\frac{1}{2}}\sigma^3+3f[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+2(\eta_2-\eta_3)'V(Y_2-Y_3)]^2/$$
$$8G^{\frac{1}{2}}\sigma^5-f[\eta_2'V^2\eta_2+2\eta_2'V(Y_2-Y_1)+(Y_2-Y_1)'(Y_2-Y_1)]/2G^{3/2}\sigma$$
$$+3f[\eta_2'V\eta_2+2\eta_2'(Y_2-Y_1)]^2/8G^{5/2}\sigma+f[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+$$
$$2(\eta_2-\eta_3)'(Y_2-Y_3)][\eta_2'V\eta_2+2\eta_2'(Y_2-Y_1)]/2G^{3/2}\sigma^3$$
$$+[-\eta_2'V(\eta_2-\eta_3)-2\eta_2'(Y_2-Y_1)+\eta_2'(Y_3-Y_1)+\eta_3'(Y_2-Y_1)] \times$$
$$[(\eta_2-\eta_3)'V(\eta_2-\eta_3)+2(\eta_2-\eta_3)'(Y_2-Y_3)]/2G^{\frac{1}{2}}\sigma^3$$

$$+[-\eta_2'V(\eta_2-\eta_3)-2\eta_2'(Y_2-Y_1)+\eta_2'(Y_3-Y_1)+\eta_3'(Y_2-Y_1)] \times$$
$$[\eta_2'V\eta_2+2\eta_2'(Y_2-Y_1)]/2G^{3/2}\sigma.$$

By going through exactly the same arguments as in Section 1.1, if $|\rho| < 1$, we have

$$(1.56)\ \mathcal{E}\hat{P}_{21}(Y_1,Y_2,Y_3,V)$$
$$= \Psi(\alpha,\beta;\rho)$$
$$+ \frac{1}{m}\Big\{\varphi_1(\alpha)\Phi_1(G^{\frac{1}{2}}q/2\tau)[G^{3/2}/32-(p-1)G^{\frac{1}{2}}/4-3(p-1)/2G^{\frac{1}{2}}]$$
$$+\varphi_1(\beta)\Phi_1(\xi/2\sigma\tau)(e/4\sigma)[(2p+1)/2+(e^2+8\tau^2+48(G+b-d)$$
$$+48(p-1))/8\sigma^2-3e^2/2\sigma^4]-\varphi_1(\alpha)\varphi_1(-G^{\frac{1}{2}}q/2\tau)(1/\tau) \times$$
$$[\tfrac{1}{2}f(G/8+3/2)+(2(G^2-d^2)+12f+ef(f+6)/\sigma^2)/16$$
$$-(G(f+6)\xi/16\tau^2)(1-f^2/G\sigma^2)]+\varphi_1(\beta)\varphi_1(-\xi/2\sigma\tau)(1/\tau) \times$$
$$\{-(f/2)[(e^2+8\tau^2+48(G+q))/8\sigma^2-3e^2/2\sigma^3]+[2(G^2-d^2)$$
$$+12(G+f)+ef(f+6)/\sigma^2]/16+(Gq/16\tau^2)[-G^2+Gd+3Gb+bd$$
$$-4d^2+6(G+3b)+2df(d+6)/G-12ef/\sigma^2$$
$$+ef^2(f+6)/G\sigma^2]\}+\varphi_2(\alpha,\beta;\rho)\{[-12-3(p-1)f/\sigma^2$$
$$-(3(p-1)f+6d)/G+f^2(f+6)/G\sigma^2]/G^{\frac{1}{2}}\sigma$$
$$+(1/2\tau^2\sigma)[f/G^{\frac{1}{2}}+G^{\frac{1}{2}}q\xi/4\tau^2][-f^2+6(b+d)-6d^2/G$$
$$-6f^2/\sigma^2+f^2(f+6)/G\sigma^2]+(G^{\frac{1}{2}}\xi/16\tau^2\sigma)(1-f/G\sigma^2)$$
$$+(G^{\frac{1}{2}}q/16\tau^2\sigma)[-G^2+Gd+3Gb+bd-4d^2+6(G+3b)$$
$$+2df(d+6)/G-12ef/\sigma^2+ef^2(f+6)/G\sigma^2]\}\Big\}+O(m^{-2}).$$

Interchanging the subscripts 2 and 3 in (1.56) gives the expansion for $\mathcal{E}\hat{P}_{31}(Y_1,Y_2,Y_3,V)$.

CHAPTER 2

ASYMPTOTIC PMC'S OF NEAREST NEIGHBOR RULES BASED ON RANKS

2.0 Introduction

The nearest neighbor (NN) rule for classifying an observation Z into one of two given populations π_1 and π_2 was first introduced and studied by Fix and Hodges (1951). The rule can be described as follows: Let $(X_1, \ldots, X_{n_1})$ and $(Y_1, \ldots, Y_{n_2})$ be random training samples from π_1 and π_2, respectively. Using a distance function d, rank the distance of all the observations from Z, and classify Z into π_i if the nearest observation to Z comes from π_i. This rule was also studied by Cover and Hart (1967), Cover (1968), who generally considered sampling from a population which is a mixture of π_1 and π_2.

The NN rule can't be applied if the observations are available only in terms of their ranks (or relative orders). In this chapter a rule is suggested which uses the basic idea of the NN rule but it is expressed only in terms of the ranks of the observations. The rule is given below and it is termed as the Modified Nearest Neighbor (MNN) rule.

Pool all the observations Z, X_i's, and Y_j's, and note their relative orders (or ranks). Let U and V be the nearest observations to Z from the left and from the right, respectively, in the pooled training sample. When either U or V is not defined we define it to be Z. The MNN rule can now be described as follows:

(i) If Z is the smallest observation, then classify it to π_1 (or π_2) when V is an X (or Y) observation.

(ii) If Z is the largest observation, classify it to π_1 (or π_2) when U is an X (or Y) observation.

(iii) If both U and V are X (or Y) observations, classify Z into π_1 (or π_2).

(iv) If U and V are not from the same population, classify Z into π_1 and π_2 with probability $\frac{1}{2}$ and $\frac{1}{2}$, respectively.

The basic idea of this rule is taken from Anderson (1966) where he discusses classification rules based on tolerance regions.

In Section 2.1 we have derived the asymptotic (as $n_1, n_2 \to \infty$) PMC of the MNN rule. It turns out that the asymptotic PMC of our rule is the same as the respective asymptotic PMC of the NN rule as derived by Fix and Hodges (1951).

Moreover, to reduce randomization in the MNN rule, we may modify the rule in the following way. If the left nearest neighbor U_1 and the right nearest neighbor V_1 are not from the same population, consider the next smaller and the next larger observations and denote the new left neighbor by U_2 and the new right neighbor by V_2. Then classify Z into π_1 (or π_2) if both U_2 and V_2 are X (or Y) observations; if U_2 and V_2 are not from the same population, classify Z into π_1 and π_2 with probability $\frac{1}{2}$ and $\frac{1}{2}$, respectively. This will be called the two-stage MNN rule. When U_2 and V_2 are not from the same population, we may consider the next smaller and the next larger observations and classify according to the new left and right neighbors U_3 and V_3 as above. This defines the three-stage MNN rule. We shall derive the asymptotic PMC's of the two-stage, and the three-stage MNN rules. When training samples are drawn from a population which is a mixture of π_1 and π_2 we derive the asymptotic risks of the two-stage and the three-stage MNN rules and extend this to obtain the asymptotic risk of the K-stage MNN rule. It is shown that this multi-stage MNN rule reduces not only the probability of randomization but also the asymptotic risk.

In section 2.4 we define the rank-analogue of the K_n-nearest neighbor (K_n-NN) rule. The K_n-NN rule was first introduced and studied by Fix and Hodges (1951), and later modified by Cover (1968). The modified rule can be described as follows. Let M_{n_i} be the number of observations in the pooled training sample from the population π_i that belong to the k_n nearest neighbors (with respect to some distance measure) of Z. Then the K_n-NN rule decides Z as π_i, if $M_{n_i} = \max_{j=1,2} M_{n_j}$. We propose a "Modified K_n-Nearest Neighbor" (MK_n-NN) rule, which uses the basic idea of the K_n-NN rule but it is expressed only in terms of the ranks of the observations. The rule is given below.

Let U_n be the $k_{n,1}$th nearest observation to Z from the left and V_n be the $k_{n,2}$th nearest observation to Z from the right in the pooled training sample. When U_n (or V_n) is not defined as described above, we define it to be the smallest (or the largest) observation in the pooled sample (including Z). Then the MK_n-NN rule is defined as follows:

(i) If there are more X (or Y) observations in the closed interval $[U_n, V_n]$, classify Z into π_1 (or π_2).

(ii) If there are equal numbers of X observations and Y observations in $[U_n, V_n]$, classify Z into π_1 and π_2 with probability $\frac{1}{2}$ and $\frac{1}{2}$, respectively.

We shall derive the asymptotic PMC of the MK_n-NN rule when $K_{n,i}\to\infty$ and $k_{n,i}/n\to 0$ as $n = \min\{n_1,n_2\}\to\infty$. When training samples are drawn from a population which is a mixture of π_1 and π_2, the asymptotic risk of the MK_n-NN rule turns out to be exactly the Bayes risk.

The K_n-NN rule was obtained using the K_n-NN estimates of the density functions as suggested by Fix and Hodges (1951) and Loftsgaarden and Quesenberry (1965). However, it may be noted that the density functions can't be estimated using the ranked observations only.

2.1 Asymptotic values of the conditional PMC of the MNN rule.

Let the c.d.f.'s of X_i and Y_j be F_1 and F_2, respectively. We shall assume that f_1 and f_2 are the p.d.f.'s corresponding to F_1 and F_2, respectively, with respect to Lebesgue measure. Denote by $X_{(1)} < X_{(2)} < \ldots < X_{(n_1)}$, $Y_{(1)} < Y_{(2)} < \ldots < Y_{(n_2)}$ the order statistics of $(X_1, \ldots, X_{n_1})$ and $(Y_1, \ldots, Y_{n_2})$, respectively.

Let $P_{n_1,n_2}(z)$ and $Q_{n_1,n_2}(z)$ be the conditional probabilities that the MNN rule classifies the observation Z into π_1 and π_2, respectively, given $Z = z$. Note that $P_{n_1,n_2}(z)$ is the conditional PCC and $Q_{n_1,n_2}(z)$ is the conditional PMC when $Z \sim F_1$, given $Z = z$. We can write

$$(2.1)\quad P_{n_1,n_2}(z) = \sum_{i=1}^{4} P_{n_1,n_2,i}(z)$$

where

$$(2.2)\quad P_{n_1,n_2,1}(z) = \Pr\{Z \leq X_{(1)} < Y_{(1)} | Z = z\},$$

$$(2.3)\quad P_{n_1,n_2,2}(z) = \Pr\{Z \geq X_{(n_1)} > Y_{(n_2)} | Z = z\},$$

$$(2.4)\quad P_{n_1,n_2,3}(z) = \Pr\{X_{(i)} \leq Z \leq X_{(i+1)} \text{ for some } i=1,\ldots,n_1-1;$$

$$Y_{(j)} \notin [X_{(i)}, X_{(i+1)}] \text{ for every } j=1,\ldots,n_2 | Z=z\},$$

(2.5) $P_{n_1,n_2,4}(z) = \frac{1}{2}\Pr\{X_{(i)} \leq Z \leq Y_{(j)}$, for some i and j but no other observations fall in $[X_{(i)}, Y_{(j)}] | Z=z\}$ + $\frac{1}{2}\Pr\{Y_{(j)} \leq Z \leq X_{(i)}$ for some i and j but no other observations fall in $[Y_{(j)}, X_{(i)}]. | Z=z\}$.

Similarly, we can write

(2.6) $$Q_{n_1,n_2}(z) = \sum_{i=1}^{4} Q_{n_1,n_2,i}(z),$$

where $Q_{n_1,n_2,i}(z)$ is obtained from $P_{n_1,n_2,i}(z)$ by interchanging X and Y. Note that $Q_{n_1,n_2,4}(z) = P_{n_1,n_2,4}(z)$.

To obtain asymptotic expressions, we shall assume that $0 < \lambda < \infty$, where

(2.7) $$\lambda = \lim n_2/n_1.$$

The cases $\lambda = 0$ and $\lambda = \infty$ can be handled easily. We shall now obtain the limiting values of $P_{n_1,n_2,i}(z)$ and $Q_{n_1,n_2,i}(z)$.

<u>Lemma 2.1.</u> (i) Either $F_1(z) > 0$ or $F_2(z) > 0$ implies $P_{n_1,n_2,1}(z) \to 0$ and $Q_{n_1,n_2,1}(z) \to 0$ as $n_1,n_2 \to \infty$.

(ii) Either $F_1(z) < 1$ or $F_2(z) < 1$ implies $P_{n_1,n_2,2}(z) \to 0$ and $Q_{n_1,n_2,2}(z) \to 0$ as $n_1,n_2 \to \infty$.

<u>Proof.</u> It is sufficient to prove the result only for $P_{n_1,n_2,1}(z)$ since the other results follow along the similar line.

$$(2.8)\quad P_{n_1,n_2,1}(z) = \Pr\{Z \leq X_{(1)} < Y_{(1)} | Z=z\}$$

$$= \int_z^\infty (1-F_2(x))^{n_2} n_1(1-F_1(x))^{n_1} dF_1(x)$$

$$\leq (1-F_2(z))^{n_2}(1-F_1(z))^{n_1} \to 0 \quad \text{as} \quad n_1,n_2 \to \infty.$$

<u>Remark.</u> The above assumptions hold a.e. if either $Z \sim F_1$ or $Z \sim F_2$.

Next we shall derive the limiting value of $P_{n_1,n_2,3}(z)$. Note that

$$(2.9)\quad P_{n_1,n_2,3}(z) =$$

$$\int_{-\infty}^{z}\int_{z}^{\infty} \{1-(F_2(y)-F_2(x))\}^{n_2} n_1(n_1-1)\{1-(F_1(y)-F_1(x))\}^{n_1-2}$$

$$dF_1(y)dF_1(x)$$

For fixed z and $0 < F_1(z) < 1$, $o < F_2(z) < 1$ define

$$(2.10)\quad H_1(y-z) = (F_1(y)-F_1(z))/(1-F_1(z)) \quad \text{for} \quad y \geq z,$$

$$(2.11)\quad H_2(z-x) = (F_1(z)-F_1(x))/F_1(z) \quad \text{for} \quad x \leq z,$$

$$(2.12)\quad K_1(y-z) = (F_2(y)-F_2(z))/(1-F_2(z)) \quad \text{for} \quad y \geq z,$$

(2.13) $K_2(z-x) = (F_2(z)-F_2(x))/F_2(z)$ for $x \leq z$.

Let $u = y-z$ and $v = z-x$ for $x \leq z \leq y$. Then we may write

$$(2.14)\quad P_{n_1,n_2,3}(z) = \int_{-\infty}^{z}\int_{z}^{\infty}\{1-(F_2(y)-F_2(z))-(F_2(z)-F_2(x))\}^{n_2} \times$$

$$n_1(n_1-1)\{1-(F_1(y)-F_1(z))-(F_1(z)-F_1(x))\}^{n_1-2}dF_1(y)dF_1(x)$$

$$= \int_{-\infty}^{z}\int_{z}^{\infty}\{1-(1-F_2(z))K_1(y-z)-F_2(z)K_2(z-x)\}^{n_2}n_1(n_1-1) \times$$

$$\{1-(1-F_1(z)H_1(y-z)-F_1(z)H_2(z-x)\}^{n_1-2}dF_1(y)dF_1(x)$$

$$= \int_{0}^{\infty}\int_{0}^{\infty}\{1-(1-F_2(z))K_1(u)-F_2(z)K_2(v)\}^{n_2}n_1(n_1-1) \times$$

$$\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v).$$

Let

(2.15) $\theta = f_2(z)/f_1(z)$.

We shall show that

$$(2.16)\quad P_{n_1,n_2,3}(z)$$

$$\doteq P^*_{n_1,n_2,3}(z)$$

$$\equiv \begin{cases} \int_0^\infty\int_0^\infty \{1-\theta H_1(u)(1-F_1(z))-\theta H_2(v)F_1(z)\}^{n_2} n_1(n_1-1)\{1-(1-F_1(z))\times \\ H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v) \text{ if } \theta \leq 1 \\ \int_0^\infty\int_0^\infty \{1-(1-F_2(z))K_1(u)-F_2(z)K_2(v)\}^{n_2} n_1(n_1-1)\{1-(1/\theta)\times \\ (1-F_2(z))K_1(u)-(1/\theta)F_2(z)K_2(v)\}^{n_1-2}(1-F_2(z))F_2(z)dK_1(u) \\ dK_2(v) \text{ if } \theta > 1 \end{cases}$$

Next we shall show that

$$P^*_{n_1,n_2,3}(z) \to (1+\lambda\theta)^{-2}$$

For the above results we need to assume that z is a continuous point of both f_1 and f_2 and $f_1(z) > 0$, $f_2(z) > 0$. Note that $P^*_{n_1,n_2,3}(z)$ is obtained from $P_{n_1,n_2,3}(z)$ after replacing $K_1(u)$ and $K_2(v)$ in the integrand by $\theta(1-F_1(z))H_1(u)/(1-F_2(z))$ and $\theta F_1(z)H_2(v)/F_2(z)$, respectively, when $\theta \leq 1$. When $\theta > 1$, $P^*_{n_1,n_2,3}(z)$ is obtained from $P_{n_1,n_2,3}(z)$ after replacing $H_1(u)$ and $H_2(v)$ in the integrand by $(1/\theta)(1-F_2(z))K_1(u)/(1-F_1(z))$ and $(1/\theta)F_2(z)K_2(v)/F_1(z)$, respectively.

We shall prove that each of $P_{n_1,n_2,3}(z)$ and $P^*_{n_1,n_2,3}(z)$ is asymptotically equivalent to the corresponding integral when the domain of integration $[0,\infty) \times [0,\infty)$ is replaced by $[0,\delta] \times [0,\delta]$ for sufficiently small $\delta > 0$. Moreover, in this domain

$$K_1(u)/H_1(u) \doteq \theta(1-F_1(z))/(1-F_2(z))$$

and

$$K_2(u)/H_2(u) \doteq \theta F_1(z)/F_2(z).$$

Let us now prove the results described above.

Lemma 2.2. If z is a continuous point of both f_1 and f_2 and $f_1(z) > 0$, $f_2(z) > 0$, then for sufficiently small $\delta > 0$ we have

$$(2.17)\quad \lim_{n_1,n_2\to\infty} [P_{n_1,n_2,3}(z)-\int_0^\delta\int_0^\delta \{1-(1-F_2(z))K_1(u)-F_2(z)K_2(v)\}^{n_2} \times$$
$$n_1(n_1-1)\ \{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)\times$$
$$dH_1(u)dH_2(v)] = 0.$$

Proof. Since H_1 and H_2 are non-decreasing functions in u and v we have

$$\int_0^\delta\int_\delta^\infty \{1-(1-F_1(z))K_1(u)-F_2(z)K_2(v)\}^{n_2} n_1(n_1-1)\{1-(1-F_1(z))\ \times$$
$$H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v)$$
$$\leq n_1(n_1-1)\{1-(1-F_1(z))H_1(\delta)\}^{n_1-2} \to 0 \qquad \text{as} \quad n_1 \to \infty,$$

because $1-(1-F_1(z))H_1(\delta) < 1$ for sufficiently small $\delta > 0$. The other two integrals $\int_\delta^\infty\int_0^\delta$ and $\int_\delta^\infty\int_\delta^\infty$ can be similarly proved to be asymptotically zero, and the proof is complete.

Remark. Note that only $n_1 \to \infty$ is required to obtain the desired result.

Before commencing the next lemma, we need to develop some useful facts.

Using the definition of a density at its point of continuity, we get

$$(2.18)\quad \lim_{u\to 0} K_1(u)/u = \lim_{u\to 0}\{(F_2(z+u)-F_2(z))/u\}/(1-F_2(z))$$
$$= f_2(z)/(1-F_2(z)),$$

$$(2.19)\quad \lim_{u\to 0} H_1(u)/u = \lim_{u\to 0} \{(F_1(z+u)-F_1(z))/u\}/(1-F_1(z))$$
$$= f_1(z)/(1-F_1(z))$$

provided $f_1(z) > 0$ and $f_2(z) > 0$.

(2.18) and (2.19) entail

$$(2.20)\quad \lim_{u\to 0} K_1(u)/H_1(u) = \theta(1-F_1(z))/(1-F_2(z)).$$

Similarly,

$$(2.21)\quad \lim_{v\to 0} K_2(v)/H_2(v) = \theta F_1(z)/F_2(z)$$

If we write

$$(2.22)\quad K_1(u) = \theta(1-F_1(z))H_1(u)/(1-F_2(z)) + R_1(u)/(1-F_2(z)),$$

(2.23) $K_2(v) = \theta F_1(z)H_2(v)/F_2(z) + R_2(v)/F_2(z)$

Then $R_1(u)$ and $R_2(v)$ have the following property:

$$\lim_{u\to 0} R_1(u)/H_1(u) = \lim_{u\to 0} R_2(v)/H_2(v) = 0,$$

which is equivalent to that for every $\epsilon > 0$, there exists a $\delta > 0$ such that

(2.24) $|R_1(u)| \leq \epsilon H_1(u)$, $|R_2(v)| \leq \epsilon H_2(v)$ whenever $|u| < \delta$, $|v| < \delta$.

Lemma 2.3. Suppose z is a continuous point of both f_1 and f_2 and $f_1(z) > 0$, $f_2(z) > 0$. If $0 < \lambda < \infty$, then

$$\lim_{n_1,n_2\to\infty} [P_{n_1,n_2,3}(z) - P^*_{n_1,n_2,3}(z)] = 0$$

Proof. We shall only prove for $\theta \leq 1$ since for the case $\theta > 1$ can be similarly proved just by switching the roles of $H_1(u)$, $H_2(v)$ with $K_1(u)$, $K_2(v)$, respectively.

Obviously Lemma 2.2 is also true for $P^*_{n_1,n_2,3}(z)$. Therefore we need only to prove that for sufficiently small $\delta > 0$

$$\lim_{n_1,n_2\to\infty} [\int_0^\delta\int_0^\delta \{1-(1-F_2(z))K_1(u)-F_2(z)K_2(v)\}^{n_2} n_1(n_1-1)\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v) -$$
$$\int_0^\delta\int_0^\delta \{1-\theta(1-F_1(z))H_1(u)-\theta F_1(z)H_2(v)\}^{n_2} n_1(n_1-1) \times$$

$$\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))\,F_1(z)\times$$

$$dH_1(u)dH_2(v)\}]=0.$$

Now using (2.22) and (2.23), we can write

$$(2.25)\ \int_0^\delta\int_0^\delta\{1-(1-F_2(z))K_1(u)-F_2(z)K_2(v)\}^{n_2}n_1(n_1-1)\{1-(1-F_1(z))\times$$

$$H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v)$$

$$=\int_0^\delta\int_0^\delta\{1-[\theta(1-F_1(z))H_1(u)+R_1(u)]-[\theta F_1(z)H_2(v)+R_2(v)]\}^{n_2}\times$$

$$n_1(n_1-1)\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))dH_1(u)$$

$$dH_2(v)$$

From the mean-value theorem, then for some $0<\gamma_1<1$, $0<\gamma_2<1$, (2.25) is

$$(2.26)\ \int_0^\delta\int_0^\delta\{1-\theta(1-F_1(z))H_1(u)-\theta F_1(z)H_2(v)\}^{n_2}n_1(n_1-1)\{1-(1-F_1(z))\times$$

$$H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v)-\int_0^\delta\int_0^\delta$$

$$R_2(v)n_2\{1-\theta(1-F_1(z))H_1(u)-\theta F_1(z)H_2(v)-\gamma_2R_2(v)\}^{n_2-1}\times$$

$$n_1(n_1-1)\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)\times$$

$$dH_1(u)dH_2(v)-\int_0^\delta\int_0^\delta R_1(u)n_2\{1-\theta(1-F_1(z))H_1(u)-\gamma_1R_1(u)-\theta F_1(z)\times$$

$$H_2(v)-R_2(v)\}^{n_2-1}\ n_1(n_1-1)\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}\times$$

$$(1-F_1(z))F_1(z)dH_1(u)dH_2(v)$$

The lemma is proved if we can show that in (2.26) the second term and the third are asymptotically zero. Denote the second term by $a_{n_1,n_2}(z)$ and the third by $b_{n_1,n_2}(z)$, then

$$|a_{n_1,n_2}(z)| \leq \int_0^\delta\int_0^\delta |R_2(v)| n_2 n_1(n_1-1)\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}\ (1-F_1(z))F_1(z)dH_1(u)dH_2(v)$$

From (2.24) for every $\epsilon > 0$, we can choose sufficiently small $\delta > 0$ such that

$$(2.27)\ |a_{n_1,n_2}(z)| \leq \epsilon \int_0^\delta\int_0^\delta H_2(v)n_2n_1(n_1-1)\{1-(1-F_1(z)H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)dH_1(u)dH_2(v) \leq \epsilon \int_0^\infty\int_0^\infty H_2(v)\times n_2n_1(n_1-1)\{1-(1-F_1(z))H_1(u)-F_1(z)H_2(v)\}^{n_1-2}(1-F_1(z))F_1(z)\times dH_1(u)dH_2(v)$$

$$= \epsilon \int_0^1\int_0^1 yn_2n_1(n_1-1)\{1-(1-F_1(z))x-F_1(z)y\}^{n_1-2}(1-F_1(z))F_1(z)\ dxdy$$

Considering the following transformation

$$s = (1-F_1(z))x + F_1(z)y$$

$$t = F_1(z)y$$

then we have

$$|a_{n_1,n_2}(z)| \le \epsilon \int_0^1\int_0^s [n_2 n_1(n_1-1)t(1-s)^{n_1-2}/F_1(z)]dtds$$

$$= (\epsilon/2F_1(z)) \int_0^1 n_2 n_1(n_1-1)s^2(1-s)^{n_1-2} ds$$

$$= (\epsilon/2F_1(z))(n_2/(n_1+1))$$

Since ϵ is arbitrary and $\lambda < \infty$, $|a_{n_1,n_2}(z)| \to 0$ is proved. Similarly $|b_{n_1,n_2}(z)| \to 0$.

Lemma 2.4. Under the assumptions of Lemma 2.3, we have

(2.28) $$\lim_{n_1,n_2\to\infty} P_{n_1,n_2,3}(z) = 1/(1+\lambda\theta)^2$$

Proof. From Lemma 2.3, it is sufficient to show that $\lim P^*_{n_1,n_2,3}(z) = 1/(1+\lambda\theta)^2$.

Let

$$s = (1-F_1(z))H_1(u)+F_1(z)H_2(v)$$

$$t = F_1(z)H_2(v)$$

Then

$$P^*_{n_1,n_2,3}(z) = \int_0^1\int_0^s (1-\theta s)^{n_2} n_1(n_1-1)(1-s)^{n_1-2} dtds$$

$$= \int_0^1 n_1(n_1-1)s(1-\theta s)^{n_2}(1-s)^{n_1-2} ds$$

Case I. $\lambda\theta < 1$.

$$P^*_{n_1,n_2,3}(z) = \int_0^1 n_1(n_1-1)s\ \{\sum_{R=0}^{n_2} \binom{n_2}{k}(-\theta)^k s^k\}(1-s)^{n_1-2} ds$$

$$= n_1(n_1-1) \sum_{k=0}^{n_2} \binom{n_2}{k}(-\theta)^k (k+1)!(n_2-1)!/(n_1+k)!$$

$$= \sum_{k=0}^{n_2} [n_1!n_2!(k+1)/(n_2-k)!(n_1+k)!](-\theta)^k$$

$$\equiv \sum_{k=0}^{n_2} (-1)^k \nu_{k,n_2}$$

This is an alternating series, and

$$\nu_{k+1,n_2}/\nu_{k,n_2} = ((n_2-k)/(n_1+k+1))((k+2)/(k+1))\theta$$

$$\leq (n_2/n_1)((k+2)/(k+1))\theta$$

Since $\lambda\theta < 1$, we can choose n_2 and k sufficiently large such that ν_{k,n_2} is decreasing in k. Moreover, $\nu_{k,n_2} \to 0$ as $k\to\infty$.

Therefore

$$\lim P^*_{n_1,n_2,3}(z) = \sum_{k=0}^{\infty} (k+1)(-\lambda\theta)^k = 1/(1+\lambda\theta)^2.$$

Case II. $\lambda\theta > 1$.

$$P^*_{n_1,n_2,3}(z) = \int_0^1 n_1(n_1-1)s(1-\theta s)^{n_2}(1-s)^{n_1-2} ds$$

$$= \int_0^\theta n_1(n_1-1)(y/\theta)(1-y)^{n_2}(1-y/\theta)^{n_1-2}(1/\theta)dy$$

$$= (1/\theta^2)\int_0^\theta n_1(n_1-1)y(1-y)^{n_2}\left\{\sum_{k=0}^{n_1-2}\binom{n_1-2}{k}(-y/\theta)^k\right\}dy$$

$$= (1/\theta^2)\sum_{k=0}^{n_1-2} n_1(n_1-1)\binom{n_1-2}{k}(-1/\theta)^k\int_0^\theta(1-y)^{n_2}y^{k+1}dy$$

Let $$R^*_{n_1,n_2,3}(z) = (1/\theta^2)\sum_{k=0}^{n_1-2} n_1(n_1-1)\binom{n_1-2}{k}(-1/\theta)^k\int_0^1(1-y)^{n_2}y^{k+1}dy$$

$$= (1/\theta^2)\sum_{k=0}^{n_1-2}[n_1!n_2!(k+1)/(n_1-k-2)!(n_2+k+2)!](-1/\theta)^k$$

Since $1/\lambda\theta < 1$, by going through the same argument as in Case I, we have

$$\lim R^*_{n_1,n_2,3}(z) = (1/\lambda^2\theta^2)\sum_{k=0}^{\infty}(k+1)(-1/\lambda\theta)^k = 1/(1+\lambda\theta)^2.$$

To complete the proof of this case, it suffices to show that $\lim\,(P^*_{n_1,n_2,3}(z)-R^*_{n_1,n_2,3}(z)) = 0.$

To see this, let

$$D_{n_1,n_2,3}(z) = \theta^2(P^*_{n_1,n_2,3}(z)-R^*_{n_1,n_2,3}(z)).$$

Then

$$|D_{n_1,n_2,3}(z)| \le \sum_{k=0}^{M} n_1(n_1-1)\binom{n_1-2}{k}(1/\theta)^k\int_\theta^1(1-y)^{n_2}y^{k+1}dy$$
$$+ \sum_{k=M}^{n_1-2} n_1(n_1-1)\binom{n_2-2}{k}(1/\theta)^k\int_0^1(1-y)^{n_2}y^{k+1}dy$$

Since $\sum_{k=0}^{n_1-2} n_1(n_1-1)\binom{n_2-2}{k}(1/\theta)^k\int_0^1(1-y)^{n_2}y^{k+1}dy$ converges, we can choose M sufficiently large so that for every given $\epsilon > 0$, the second term of right side of the above inequality is dominated by ϵ. Hence for every $\epsilon > 0$, there exists M such that

$$\lim |D_{n_1,n_2,3}(z)| \le \lim \sum_{k=0}^{M} n_1(n_1-1)\binom{n_1-2}{k}(1/\theta)^k \int_\theta^1 (1-y)^{n_2} \times$$

$$y^{k+1}dy + \epsilon \le \lim \sum_{k=0}^{M} n_1(n_1-1)\binom{n_1-2}{k}(1/\theta)^k(1-\theta)^{n_2+1} + \epsilon = \epsilon$$

Thus $\lim D_{n_1,n_2,3}(z) = 0$ since ϵ is arbitrary.

Case III. $\lambda\theta = 1$.

We have $\lim P^*_{n_1,n_2,3}(z) = \sum_{k=0}^{\infty} (k+1)(-1)^k$

Then by the method of Abel of summability (see Widder (1961), p. 309-313), we get

$$\lim P^*_{n_1,n_2,3}(z) = \lim_{x\to 1^-} \sum_{k=0}^{\infty} (k+1)(-x)^k$$

$$= \lim_{x\to 1^-} 1/(1+x)^2 = 1/4 = 1/(1+\lambda\theta)^2$$

Now the proof of Lemma 2.4 is complete.

Lemma 2.5. Under the assumptions of Lemma 2.3, we have

(2.29) $\lim_{n_1,n_2\to\infty} P_{n_1,n_2,4}(z) = \lambda\theta/(1+\lambda\theta)^2.$

Proof. Going through the same arguments as we did for $P_{n_1,n_2,3}(z)$ (Lemma 2.2 through Lemma 2.4), we have

$$P_{n_1,n_2,4}(z) = \tfrac{1}{2}\int_{-\infty}^{z}\int_{z}^{\infty} n_1n_2\{1-(F_2(y)-F_2(x))\}^{n_2-1}\{1-(F_1(y)$$

$$-F_1(x))\}^{n_1-1}(dF_2(y)dF_1(x)+dF_1(y)dF_2(x))$$

$$= \tfrac{1}{2}\int_0^{\infty}\int_0^{\infty} n_1n_2\{1-(1-F_2(z))K_1(u)-F_2(z)K_2(v)\}^{n_2-1} \times$$

$$\{1-(1-F_1(z)H_1(u)-F_1(z)H_2(v)\}^{n_1-1} \times$$

$$\{(1-F_2(z))F_1(z)dK_1(u)dH_2(v)+(1-F_1(z))F_2(z)\times$$

$$dH_1(u)dK_2(v)\}$$

$$\doteq \theta(n_2/n_1+1))\int_0^{\infty}\int_0^{\infty}(n_1+1)n_1\{1-\theta(1-F_1(z))H_1(u)-$$

$$\theta F_1(z)H_2(v)\}^{n_2-1}\{1-(1-F_1(z))H_1(u)-F_1(z)\times$$

$$H_2(v)\}^{n_1-1}(1-F_1(z))F_1(z)dH_1(u)dH_2(v)$$

$$\to \theta\lambda(1/(1+\lambda\theta)^2) = \lambda\theta/(1+\lambda\theta)^2 \quad \text{as} \quad n_1,n_2 \to \infty.$$

Combining all the previous results, we have the following theorem.

Theorem 2.1. Suppose z is a continuous point of both f_1 and f_2 and $f_1(z) > 0$, $f_2(z) > 0$. If $0 < \lambda < \infty$, then

$$(2.30)\quad \lim_{n_1,n_2\to\infty} P_{n_1,n_2}(z) = f_1(z)/(f_1(z)+\lambda f_2(z)),\quad \text{and}$$

$$\lim_{n_1,n_2\to\infty} Q_{n_1,n_2}(z) = \lambda f_2(z)/(f_1(z)+\lambda f_2(z)).$$

<u>Proof.</u> From Lemma 2.1, Lemma 2.4, and Lemma 2.5, we have

$$\lim P_{n_1,n_2}(z) = \lim \sum_{i=1}^{4} P_{n_1,n_2,i}(z)$$

$$= 1/(1+\lambda\theta)^2 + \lambda\theta/(1+\lambda\theta)^2 = 1/(1+\lambda\theta)$$

$$= f_1(z)/(f_1(z)+\lambda f_2(z))$$

Let $\pi_{n_1,n_2,1}$ and $\pi_{n_1,n_2,2}$ be the PMC as Z coming from π_1 or π_2, respectively. If $\Pr\{f_i(Z) > 0,\ f_i$ is continuous at $Z|\pi_j\} = 1$, i=1,2., j=1,2., the the followings are immediate consequences of Theorem 2.1.

$$(2.31)\quad \lim_{n_1,n_2\to\infty} \pi_{n_1,n_2,1} = \int[\lambda f_2(z)/(f_1(z)+\lambda f_2(z)]f_1(z)dz$$

$$= \int[\lambda f_1(z)f_2(z)/(f_1(z)+\lambda f_2(z))]dz$$

and

$$(2.32)\quad \lim_{n_1,n_2\to\infty} \pi_{n_1,n_2,2} = \int[f_1(z)f_2(z)/(f_1(z)+\lambda f_2(z))]dz$$

Suppose the training samples are drawn from a population which is a mixture of Π_1 and Π_2 in the proportion ξ_1 and ξ_2. Then $\lambda = \xi_2/\xi_1$, and the asymptotic risk of the MNN rule (assuming cost from misclassification is 1) is

$$(2.33)\ R = \xi_1\int[(\xi_2/\xi_1)f_1(z)f_2(z)/(f_1(z)+(\xi_2/\xi_1)f_2(z))]dz$$

$$+ \xi_2\int[f_1(z)f_2(z)/(f_1(z)+(\xi_2/\xi_1)f_2(z))]dz$$

$$= 2\int[\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+\xi_2f_2(z))]dz$$

$$\leq 2\int\min\{\xi_1f_1(z),\ \xi_2f_2(z)\}dz$$

$$= 2R^*,$$

where R^* is the Bayes risk with respect to to prior probabilities ξ_1 and ξ_2; namely, the asymptotic probability of error of the MNN rule is bounded above by twice of the Bayes probability of error.

2.2 An alternative approach to obtain the asymptotic conditional PMC of the MNN rule.

Let U be the left nearest neighbor of Z and V the right nearest neighbor of Z. Then the conditional probability of classifying Z into Π_1, given $Z = z$, $U = u$, $V = v$, is

$$(2.34)\ P_{n_1,n_2}(z,u,v) = A(n_1,n_2)/B(n_1,n_2),$$

where

$$A(n_1,n_2) = C_1(n_1,n_2)+C_3(n_1,n_2)+C_5(n_1,n_2)+\tfrac{1}{2}C_7(n_1,n_2),$$

$$B(n_1,n_2) = \sum_{i=1}^{7} C_i(n_1,n_2)$$

and

$$C_1(n_1,n_2) = n_1[1-F_1(v)]^{n_1-1}[1-F_2(v)]^{n_2}f_1(v)$$

$$C_2(n_1,n_2) = n_2[1-F_2(v)]^{n_2-1}[1-F_1(v)]^{n_1}f_2(v)$$

$$C_3(n_1,n_2) = n_1[F_1(u)]^{n_1-1}[F_2(u)]^{n_2}f_1(u)$$

$$C_4(n_1,n_2) = n_2[F_2(u)]^{n_2-1}[F_1(u)]^{n_1}f_2(u)$$

$$C_5(n_1,n_2) = n_1(n_1-1)[1-(F_1(v)-F_1(u))]^{n_1-2}[1-(F_2(v)-F_2(u))]^{n_2}\times f_1(u)f_1(v)$$

$$C_6(n_1,n_2) = n_2(n_2-1)[1-(F_2(v)-F_2(u))]^{n_2-2}[1-(F_1(v)-F_1(u))]^{n_1}\times f_2(u)f_2(v)$$

$$C_7(n_1,n_2) = n_1n_2[1-(F_1(v)-F_1(u))]^{n_1-1}[1-(F_2(v)-F_2(u))]^{n_2-1}\times [f_1(v)f_2(u)+f_1(u)f_2(v)],$$

here $C_1(n_1,n_2)$, $C_2(n_1,n_2)$, $C_3(n_1,n_2)$, $C_4(n_1,n_2)$, $C_5(n_1,n_2)$, $C_6(n_1,n_2)$, and $C_7(n_1,n_2)$ divided by the conditional joint density of U and V given $Z = z$ are the conditional probabilities of the events $\{Z \le X_{(1)} < Y_{(1)}\}$, $\{Z \le Y_{(1)} < X_{(1)}\}$, $\{Z \ge X_{(n_1)} > Y_{(n_2)}\}$, $\{Z \ge Y_{(n_2)} > X_{(n_1)}\}$, $\{X_{(i)} \le Z \le X_{(i+1)}$, for some $i=1,\ldots,n_1-1$; and no Y_j's fall in $[X_{(i)}, X_{(i+1)}]\}$, $\{Y_{(j)} \le Z \le Y_{(j+1)}$, for some $j=1,\ldots,n_2-1$; and no X_i's fall in $[Y_{(j)}, Y_{(j+1)}]\}$, $\{X_{(i)} \le Z \le Y_{(j)}$, for some i and j and no other observations fall in $[X_{(i)}, Y_{(j)}]$; or $Y_{(j)} \le Z \le X_{(i)}$ for some i and j and no other observations fall in $[Y_{(j)}, X_{(i)}]\}$, respectively, given $Z = z$, $U = u$, and $V = v$.

We begin with the following lemma.

<u>Lemma 2.6</u> Either f_1 is continuous at z and $f_1(z) > 0$ or f_2 is continuous at z and $f_2(z) > 0$ implies that U and V converge to z in probability as $n_1, n_2 \to \infty$.

<u>Proof.</u> By symmetry, it suffices to show that U converges to z in probability.

For every sufficiently small $\epsilon > 0$

$$\Pr\{Z-U > \epsilon | Z=z\} = \Pr\{U < Z-\epsilon | Z=z\}$$

$$= \{1-(F_1(z)-F_1(z-\epsilon))\}^{n_1}\{1-(F_2(z)-F_2(z-\epsilon))\}^{n_2} \to 0$$

as $n_1, n_2 \to \infty$ since either $1-(F_1(z)-F_1(z-\epsilon)) < 1$ or $1-(F_2(z)-F_2(z-\epsilon)) < 1$.

<u>An alternative proof of Theorem 2.1:</u>

First we can easily see that $C_1(n_1,n_2)$, $C_2(n_1,n_2)$, $C_3(n_1,n_2)$

and $C_4(n_1,n_2)$ converge to zero in probability as $n_1,n_2 \to \infty$ since $0 < F_1(z) < 1$, $0 < F_2(z) < 1$, U and V converge to z in probability (Lemma 2.6), and the density functions are continuous.

Thus

$$\text{plim}(A(n_1,n_2)/B(n_1,n_2))$$

$$= \text{plim}(C_5(n_1,n_2)+\tfrac{1}{2}C_7(n_1,n_2))/[C_5(n_1,n_2)+C_6(n_1,n_2)$$

$$+C_7(n_1,n_2)].$$

We can write

(2.35) $$C_5(n_1,n_2)+\tfrac{1}{2}C_7(n_1,n_2) = n_1^2[1-(F_1(v)-F_1(u))]^{n_1-2}[1-(F_2(v)-F_2(u))]^{n_2-2}\{((n_1-1)/n_1)[1-(F_2(v)-F_2(u))]^2f_1(u)f_1(v)$$
$$+\tfrac{1}{2}(n_2/n_1)[1-(F_1(v)-F_1(u))][1-(F_2(v)-F_2(u))][f_1(v)f_2(u)$$
$$+f_1(u)f_2(v)]\}$$

and

(2.36) $$C_5(n_1,n_2)+C_6(n_1,n_2)+C_7(n_1,n_2) = n_1^2[1-(F_1(v)-F_1(u)]^{n_1-2}$$
$$[1-(F_2(v)-F_2(u)]^{n_2-2}\{((n_1-1)/n_1)[1-(F_2(v)-F_2(u))]^2f_1(u) \times$$

$$f_1(v)+(n_2(n_2-1)/n_1^2)[1-(F_1(v)-F_1(u))]^2 f_2(u)f_2(v)+(n_2/n_1) \times$$

$$[1-(F_1(v)-F_1(u))][1-(F_2(v)-F_2(u))][f_1(v)f_2(u)+f_1(u)f_2(v)]\}$$

Hence, by the same reasons as we stated above, we have

$$(2.37)\quad \mathrm{p\,lim}(A(n_1,n_2)/B(n_1,n_2)) = (f_1^2(z)+\lambda f_1(z)f_2(z))/(f_1^2(z)+$$

$$\lambda^2 f_2^2(z)+2\lambda f_1(z)f_2(z)) = f_1(z)/(f_1(z)+\lambda f_2(z));$$

namely,

$$(2.38)\quad \mathrm{p\,lim}\ P_{n_1,n_2}(z,u,v) = f_1(z)/(f_1(z)+\lambda f_2(z))$$

Therefore, by the Lebesgue dominated convergence theorem, we obtain

$$\begin{aligned}\lim P_{n_1,n_2}(z) &= \lim \mathcal{E}\, P_{n_1,n_2}(z,U,V)\\ &= \mathcal{E} \lim P_{n_1,n_2}(z,U,V)\\ &= \mathcal{E}[f_1(z)/(f_1(z)+\lambda f_2(z)]\\ &= f_1(z)/(f_1(z)+\lambda f_2(z)).\end{aligned}$$

Also,

$$\lim Q_{n_1,n_2}(z) = \lim(1-P_{n_1,n_2}(z)) = \lambda f_2(z)/(f_1(z)+\lambda f_2(z)).$$

The proof is now complete.

2.3 The asymptotic conditional PMC of the multi-stage MNN rule.

The following lemma leads us to assume without loss of generality (with probability one) that the k left nearest neighbors $U_1, U_2, \ldots, U_k$ and the k right nearest neighbors $V_1, V_2, \ldots, V_k$ are well-defined in the K-stage MNN rule. Let

(2.39) $n = \min\{n_1, n_2\}$.

Lemma 2.7. If $k/n \to 0$ as $n \to \infty$, then

(i) Pr{There are at least k observations to the right of z for sufficiently large n} = 1,

(ii) Pr{There are at least k observations to the left of z for sufficiently large n} = 1.

Proof. We shall prove (i). Since continuity of distribution functions is assumed and it is known that either $Z \sim F_1$ or $Z \sim F_2$, it is then true with probability one that either $0 < F_1(z) < 1$ or $0 < F_2(z) < 1$. Suppose $0 < F_1(z) < 1$, and define

$$W_i = X_{iz,\infty,}(X_i) \qquad i = 1, \ldots, n_1.$$

Then $\mathcal{E}(W_1) = 1-F_1(z) > 0$. By the strong law of large numbers, we have

$$\Pr\{(1/n_1)\sum_{i=1}^{n_1} W_i \to \mathcal{E}(W_1) > 0 \text{ as } n_1 \to \infty\} = 1.$$

Now since $k/n_1 \to 0$ as $n_1 \to \infty$ and $\mathcal{E}(W_1) > 0$, there exists an integer N such that

$$k/n_1 \le \tfrac{1}{2}\mathcal{E}(W_1) \quad \text{for} \quad n_1 \ge N.$$

Consequently,

$$\Pr\{\sum_{i=1}^{n_1} W_i \ge k \quad \text{for} \quad n \quad \text{sufficiently large}\} = 1,$$

which completes the proof.

Let $P^{(k)}_{n_1,n_2}(z;u_1,\ldots,u_k,v_1,\ldots,v_k)$ (or $Q^{(k)}_{n_1,n_2}(z;u_1,\ldots,u_k,$ $v_1,\ldots,v_k)$) be the conditional probability of the K-stage MNN rule classifying Z into π_1 (or π_2), given $Z=z$, $U_i=u_i$, $V_i=v_i$ for $i = 1, \ldots, k$. Let $P^{(k)}_{n_1,n_2}(z)$ (or $Q^{(k)}_{n_1,n_2}(z)$) be the conditional probability of the K-stage MNN rule classifying Z into π_1 (or π_2), given $Z=z$. The limiting value of $P^{(k)}_{n_1,n_2}(z)$ is obtained through the limiting value of $P^{(k)}_{n_1,n_2}(z);u_1,\ldots,u_k;$ $v_1,\ldots,v_k)$ by using the following lemma.

<u>Lemma 2.8.</u> Suppose either f_1 is continuous at z with $f_1(z)>0$ or f_2 is continuous at z with $f_2(z)>0$. If $k/n\to 0$ as $n\to\infty$, then $U_j\to z$, $V_j\to z$ in probability as $n\to\infty$ for $j=1,\ldots,k$.

<u>Proof.</u> We shall only prove that $U_k\to z$ in probability. Suppose f_1 is continuous at z and $f_1(z)>0$, then for every sufficiently small $\epsilon > 0$,

$$\Pr\{Z-U_k > \epsilon | Z=z\}$$

$$\le \Pr\{\text{There are at most } (k-1) \text{ observations lying in the}$$

interval $(z-\varepsilon, z)$.}

$$= \sum_{i=0}^{k-1} \binom{n_1}{i} q^i (1-q)^{n_1-i} .$$

$$= \Pr\{W_{n_1}/n_1 \leq k/n_1 | Z=z\}$$

where W_{n_1} is the number of X observations lying in the interval $(z-\varepsilon, z)$.

Since by the law of large numbers,

$$W_{n_1}/n_1 \overset{a.s.}{\to} F_1(z)-F_1(z-\varepsilon) > 0 \quad \text{and}$$

$k/n_1 \to 0$, we immediately have

$$\Pr\{Z-U_k > \varepsilon | Z=Z\} \to 0 \quad \text{as} \quad n \to \infty$$

for every sufficiently small $\varepsilon > 0$, which completes the proof.

<u>Two-stage MNN rule.</u>

Define

(2.40) $D_1(n_1,n_2)$

$$= n_1(n_1-1)(n_1-2)(n_1-3)[1-(F_1(v_2)-F_1(u_2)]^{n_1-4}[1-(F_2(v_2)$$

$$-F_2(u_2))]^{n_2} f_1(u_1)f_1(v_1)f_1(u_2)f_1(v_2)+n_1(n_1-1)(n_1-2)n_2 \times$$

$$[1-(F_1(v_2)-F_1(u_2))]^{n_1-3}[1-(F_2(v_2)-F_2(u_2))]^{n_2-1} \times$$

$$[f_1(u_1)f_1(v_1)f_1(u_2)f_2(v_2)+f_1(u_1)f_1(v_1)f_2(u_2)f_1(v_2)+$$

$$f_1(u_1)f_2(v_1)f_1(u_2)f_1(v_2)+f_2(u_1)f_1(v_1)f_1(u_2)f_1(v_2)]+$$

$$n_1(n_1-1)n_2(n_2-1)[1-(F_1(v_2)-F_1(u_2))^{n_1-2}[1-(F_2(v_2)$$

$$-F_1(u_2))]^{n_2-2} f_1(u_1)f_1(v_1)f_2(u_2)f_2(v_2),$$

(2.41) $D_2(n_1,n_2)$

$$= n_2(n_2-1)(n_2-2)(n_2-3)[1-(F_1(v_2)-F_1(u_2))]^{n_1}[1-(F_2(v_2)$$

$$- F_2(u_2))]^{n_2-4} f_2(u_1)f_2(v_1)f_2(u_2)f_2(v_2)+n_1n_2(n_2-1)(n_2-2)$$

$$[1-(F_1(v_2)-F_1(u_2))]^{n_1-1}[1-(F_2(v_2)-F_2(u_2))]^{n_2-3}$$

$$[f_2(u_1)f_2(v_1)f_2(u_2)f_1(v_2)+f_2(u_1)f_2(v_1)f_1(u_2)f_2(v_2)+$$

$$f_2(u_1)f_1(v_1)f_2(u_2)f_2(v_2)+f_1(u_1)f_2(v_1)f_2(u_2)f_2(v_2)]+$$

$$n_1(n_1-1)(n_2)(n_2-1)[1-(F_1(v_2)-F_1(u_2))]^{n_1-2}[1-(F_2(v_2)$$

$$-F_2(u_2))]^{n_2-1} f_2(u_1)f_2(v_1)f_1(u_2)f_1(v_2),$$

$$(2.42)\quad D_3(n_1,n_2)$$

$$=n_1(n_1-1)n_2(n_2-1)[1-(F_1(v_2)-F_1(u_2))]^{n_1-2}[1-(F_2(v_2)$$

$$-F_2(u_2))]^{n_2-2}[f_1(u_1)f_2(v_1)f_1(u_2)f_2(v_2)+f_1(u_1)f_2(v_1)f_2(u_2)\times$$

$$f_1(v_2)+f_2(u_1)f_1(v_1)f_1(u_2)f_2(v_2)+f_2(u_1)f_1(v_1)f_2(u_2)f_1(v_2)],$$

where $D_1(n_1,n_2)$, $D_2(n_1,n_2)$, $D_3(n_1,n_2)$ are respectively proportional to the conditional probabilities of classifying Z into Π_1, classifying Z into Π_2, and randomization, given $Z=z$, $U_i=u_i$, $V_i=v_i$, $i = 1, 2$. And the configurations are {XXZXX, or XXZXY, or YXZXX, or XXZYX, or XYZXX, or YXZXY}, {YYZYY, or YYZYX, or XYZYY, or YYZXY, or YXZYY, or XYZYX}, and {XXZYY, or YXZYX, or XYZXY, or YYZXX}, respectively. Then using Lemma 2.7, we have

$$(2.43)\quad P^{(2)}_{n_1,n_2}(z;u_1,u_2,v_1,v_2)$$

$$\underset{\text{a.s.}}{\doteq}(D_1(n_1,n_2)+\tfrac{1}{2}D_3(n_1,n_2))/(D_1(n_1,n_2)+D_2(n_1,n_2)+D_3(n_1,n_2))$$

In the same manner as in section 2.2, we get

$$(2.44)\quad \underset{n\to\infty}{\text{plim}}\ P^{(2)}_{n_1,n_2}(z;u_1,u_2,v_1,v_2)$$

$$=(f_1^4(z)+4\lambda f_1^3(z)f_2(z)+\lambda^2f_1^2(z)f_2^2(z)+2\lambda^2f_1^2(z)f_2^2(z))/(f_1^4(z)$$

$$+4\lambda f_1^3(z)f_2(z)+\lambda^2f_1^2(z)f_2^2(z)+\lambda^4f_2^4(z)+4\lambda^3f_1(z)f_2^3(z)$$

$$+\lambda^2f_1^2(z)f_2^2(z)+4\lambda^2f_1^2(z)f_2^2(z))$$

$$= f_1^2(z)(f_1(z)+3\lambda f_2(z))(f_1(z)+\lambda f_2(z))/(f_1(z)+\lambda f_2(z))^4$$

$$= f_1^2(z)(f_1(z)+3\lambda f_2(z))/(f_1(z)+\lambda f_2(z))^3$$

Thus, by the dominated convergence theorem, we get

$$(2.45)\ \lim_{n\to\infty} P^{(2)}_{n_1,n_2}(z)$$

$$= f_1^2(z)(f_1(z)+3\lambda f_2(z))/(f_1(z)+\lambda f_2(z))^3$$

and

$$(2.46)\ \lim_{n\to\infty} Q^{(2)}_{n_1,n_2}(z)$$

$$= f_2^2(z)(f_2(z)+3\lambda^2 f_1(z))/(f_1(z)+\lambda f_2(z))^3$$

Similarly, the asymptotic conditional probability of randomization (or tie), given $Z=z$ is

$$(2.47)\ \lim_{n\to\infty} T^{(2)}_{n_1,n_2}(z)$$

$$= 4\lambda^2 f_1^2(z)f_2^2(z)/(f_1(z)+\lambda f_2(z))^4$$

which is exactly the square of the asymptotic conditional probability of randomization of the MNN rule. (Recall that the asymptotic conditional probability of randomization of the MNN rule (see 2.29) is $\lim_{n\to\infty}(P_{n_1,n_2,4}(z)+Q_{n_1,n_2,4}(z) = \lim_{n\to\infty} 2P_{n_1,n_3,4}(z) = 2\lambda f_1(z)f_2(z)/(f_1(z)+\lambda f_2(z))^2$.)

When the training samples are drawn from a population which is a mixture of Π_1 and Π_2 with prior probabilities ξ_1 and ξ_2, then $\lambda = \xi_2/\xi_1$. If $\Pr\{f_1(Z) > 0$, f_1 is continuous at $Z|\Pi_j\} = 1$, $i, j = 1, 2$, then the asymptotic risk of this rule is

(2.48) $R^{(2)}$

$$= \int [\xi_2 f_2(z) f_1^2(z)(f_1(z)+3\lambda f_2(z))/(f_1(z)+\lambda f_2(z))^3]dz$$

$$+\int [\xi_1 f_1(z) f_2^2(z)(f_2(z)+3\lambda^2 f_1(z))/(f_1(z)+\lambda f_2(z))^3]dz$$

$$= \int [\xi_1\xi_2 f_1(z) f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))][(\xi_1^2 f_1^2(z)+6\xi_1\xi_2 f_1(z)$$

$$f_2(z)+\xi_2^2 f_2^2(z))/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]dz$$

$$= \int [\xi_1\xi_2 f_1(z) f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))] \quad \{1+2[2\xi_1\xi_2 f_1(z) \times$$

$$f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]\}dz$$

Comparing $R^{(2)}$ with the asymptotic risk R (2.33) of the MNN rule, we have

(2.49) $R - R^{(2)}$

$$= \int [\xi_1\xi_2 f_1(z) f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))][2-(\xi_1^2 f_1^2(z)+6\xi_1\xi_2 f_1(z)$$

$$\times f_2(z)+\xi_2^2 f_2(z))/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]dz$$

$$= \int [\xi_1\xi_2 f_1(z) f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))] \quad [(\xi_1 f_1(z)-\xi_2 f_2(z))^2/$$

$$(\xi_1 f_1(z)+\xi_2 f_2(z))^2]dz \geq 0.$$

Namely, the asymptotic risk of the two-stage MNN rule is improved over that of the MNN rule unless $\xi_1 f_1(Z) = \xi_2 f_2(Z)$ a.e.

Three-stage MNN rule.

We shall omit the details. Proceeding as before, we get

(2.50) $$\lim_{n\to\infty} P^{(3)}_{n_1,n_2}(z)$$

$$= (f_1^6(z)+6\lambda f_1^5(z)f_2(z)+14\lambda^2 f_1^4(z)f_2^2(z)+10\lambda^3 f_1^3(z)f_2^3(z)$$
$$+\lambda^4 f_1^2(z)f_2^4(z))/(f_1(z)+\lambda f_2(z))^6,$$

(2.51) $$\lim_{n\to\infty} Q^{(3)}_{n_1,n_2}(z)$$

$$= (\lambda^6 f_2^6(z)+6\lambda^5 f_1(z)f_2^5(z)+14\lambda^4 f_1^2(z)f_2^4(z)+10\lambda^3 f_1^3(z)f_2^3(z)$$
$$+\lambda^2 f_1^4(z)f_2^2(z))/(f_1(z)+\lambda f_2(z))^6,$$

and the asymptotic conditional probability of randomization is

(2.52) $$\lim_{n\to\infty} T^{(3)}_{n_1,n_2}(z)$$

$$= 8\lambda^3 f_1^3(z)f_2^3(z)/(f_1(z)+\lambda f_2(z))^6,$$

which is the cube of the corresponding probability of the MNN rule. The asymptotic risk of the three-stage MNN rule is

(2.53) $$R^{(3)}$$

$$= \int[\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))][(\xi_1^5 f_1^5(z)+6\xi_1^4\xi_2 f_1^4(z)\times$$
$$f_2(z)+14\xi_1^3\xi_2^2 f_1^3(z)f_2^2(z)+10\xi_1^2\xi_2^3 f_1^2(z)f_2^3(z)+\xi_1\xi_2^4 f_1(z)f_2^4(z)$$
$$+\xi_2^5 f_2^5(z)+6\xi_1\xi_2^4 f_1(z)f_2^4(z)+14\xi_1^2\xi_2^3 f_1^2(z)f_2^3(z)$$

$$+10\xi_1^3\xi_2^2f_1^3(z)f_2^2(z)+\xi_1^4\xi_2f_1^4(z)f_2(z))/(\xi_1f_1(z)+\xi_2f_2(z))^5]dz$$

$$=\int[\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+\xi_2f_2(z))][(\xi_1^5f_1^5(z)+7\xi_1^4\xi_2f_1^4(z)$$

$$f_2(z)+24\xi_1^3\xi_2^2f_1^3(z)f_2^2(z)+24\xi_1^2\xi_2^3f_1^2(z)f_2^3(z)+7\xi_1\xi_2^4f_1(z)f_2^4(z)$$

$$+\xi_2^5f_2^5(z))/(\xi_1f_1(z)+\xi_2f_2(z))^5]$$

$$=\int[\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+\xi_2f_2(z))]\ \{1+[2\xi_1\xi_2f_1(z)f_2(z)/$$

$$(\xi_1f_1(z)+\xi_2f_2(z))^5][\xi_1^3f_1^3(z)+7\xi_1^2\xi_2f_1^2(z)f_2(z)+7\xi_1\xi_2^2f_1(z)$$

$$f_2^2(z)+\xi_2^3f_2^3(z)]\}dz$$

$$=\int[\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+\xi_2f_2(z))]\ \{1+[2\xi_1\xi_2f_1(z)f_2(z)/$$

$$(\xi_1f_1(z)+\xi_2f_2(z))^5][(\xi_1f_1(z)+\xi_2f_2(z))^3+4\xi_1\xi_2f_1(z)f_2(z)\ \times$$

$$(\xi_1f_1(z)+\xi_2f_2(z))]dz$$

$$=\int[\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+\xi_2f_2(z))]\{1+2\xi_1\xi_2f_1(z)f_2(z)/$$

$$(\xi_1f_1(z)+\xi_2f_2(z))^2+2[2\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+$$

$$\xi_2f_2(z))^2]^2\}dz$$

and

(2.54) $R^{(2)}-R^{(3)}$

$$=\int[\xi_1\xi_2f_1(z)f\ (z)/(\xi_1f_1(z)+\xi_2f_2(z))]\{2\xi_1\xi_2f_1(z)f_2(z)/$$

$$(\xi_1f_1(z)+\xi_2f_2(z))^2-2[2\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)$$

$$+\xi_2f_2(z))]^2\}dz$$

$$=\int[\xi_1\xi_2f_1(z)f_2(z)/(\xi_1f_1(z)+\xi_2f_2(z))][2\xi_1\xi_2f_1(z)f_2(z)/$$

$$(\xi_1 f_1(z)+\xi_2 f_2(z))^2] \, [(\xi_1 f_1(z)-\xi_2 f_2(z))^2/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]dz \geq 0.$$

We have computed the asymptotic risk for $k = 4$ and studied the results for different k. It appears that the asymptotic risk of the K-stage MNN rule is

(2.55) $R^{(k)}$

$$= \int [\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z)] \, \{\sum_{j=0}^{k-2} [2\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]^j + 2[2\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]^{k-1}\}dz$$

and

(2.56) $R^{(k)} - R^{(k-1)}$

$$= \int [\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))] \, [2\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]^{k-2} \, [(\xi_1 f_1(z)-\xi_2 f_2(z))^2/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]dz \geq 0.$$

Now we see that the asymptotic risk is reduced at every stage unless $\xi_1 f_1(Z) = \xi_2 f_2(Z)$, a.e., and the rate is decreasing, and the asymptotic conditional probability of randomization at the kth stage is

(2.57) $\lim_{n\to\infty} T^{(k)}_{n_1,n_2}(z)$

$$= [2\lambda f_1(z)f_2(z)/(f_1(z)+\lambda f_2(z))^2]^k$$

$$= [2\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))^2]^k.$$

Suppose we are trying to eliminate randomization at all, then the asymptotic risk is found from (2.55) to be

(2.58) $R^{(\infty)}$

$$= \int [\xi_1\xi_2 f_1(z)f_2(z)/(\xi_1 f_1(z)+\xi_2 f_2(z))] \times [(\xi_1 f_1(z)$$

$$+\xi_2 f_2(z))^2/(\xi_1^2 f_1^2(z)+\xi_2^2 f_2^2(z))]dz$$

$$= \int \{[(\xi_1^2 f_1^2(z))\ \xi_2 f_2(z)+ (\xi_2^2 f_2(z))\xi_1 f_1(z)]/(\xi_1^2 f_1^2(z)$$

$$+\xi_2^2 f_2^2(z))\}dz$$

The multi-stage MNN rule can reduce (or eliminate) randomization and reduce the asymptotic risk, but unfortunately the Bayes risk can not be attained by the rule asymptotically.

2.4 The asymptotic conditional PMC of the MK_n-NN rule.

We shall obtain the asymptotic conditional PMC when $k_{n,i}/n \to 0$ as $n \to \infty$. According to Lemma 2.7, we can assume without loss of generality (with probability one) that U_n (the $k_{n,1}$th left nearest neighbor of Z) and V_n (the $k_{n,2}$th right nearest neighbor of Z) are well defined. To avoid randomization, we set $k_{n,1} + k_{n,2} = 2k_n+1$.

Define

(2.59) $h_1(j;n)$

$$= n_1(n_1-1)\binom{n_1-2}{j}\binom{n_2}{2k_n-1-j}[F_1(v)-F_1(u)]^j$$
$$[1-(F_1(v)-F_1(u)]^{n_1-2-j}\ [F_2(v)-F_2(u)]^{2k_n-1-j}$$
$$[1-(F_2(v)-F_2(u))]^{n_2-2k_n+1+j}\ f_1(u)f_1(v),$$

(2.60) $h_2(j;n)$

$$= n_1n_2\binom{n_1-1}{j}\binom{n_2-1}{2k_n-1-j}[F_1(v)-F_1(u)]^j$$
$$[1-(F_1(v)-F_1(u))]^{n_1-1-j}\ [F_2(v)-F_2(u)]^{2k_n-1-j}$$
$$[1-(F_2(v)-F_2(u))]^{n_2-2k_n+j}[f_1(u)f_2(v)+f_1(v)f_2(u)],$$

(2.61) $h_3(j;n)$

$$= n_2(n_2-1)\binom{n_1}{j}\binom{n_2-2}{2k_n-1-j}[F_1(v)-F_1(u)]^j$$
$$[1-(F_1(v)-F_1(u))]^{n_1-j}[F_2(v)-F_2(u)]^{2k_n-1-j}$$
$$[1-(F_2(v)-F_2(u))]^{n_2-2k_n-1+j}\ f_2(u)f_2(v),$$

and

$$(2.62)\quad a_{n,1} = \sum_{j=k_n-1}^{2k_n-1} h_1(j;n)$$

$$b_{n,1} = \sum_{j=0}^{k_n-2} h_1(j;n),$$

$$(2.63)\quad a_{n,2} = \sum_{j=kn}^{2k_n-1} h_2(j;n),$$

$$b_{n,2} = \sum_{j=0}^{k_n-1} h_2(j;n),$$

$$(2.64)\quad a_{n,3} = \sum_{j=k_n+1}^{2k_n-1} h_3(j;n),$$

$$b_{n,3} = \sum_{j=0}^{k_n} h_3(j;n),$$

where $a_{n,1}$ (or $b_{n,1}$), $a_{n,2}$ (or $b_{n,2}$), $a_{n,3}$ (or $b_{n,3}$) are respectively proportional to the conditional probabilities of classifying Z into Π_1 (or Π_2) when both U_n and V_n are X observations, when only one of U_n and V_n is an X observation, when both U_n and V_n are Y observations, given $Z=z$, $U_n=u$, $V_n=v$.

Let $P_{n_1,n_2}(z;u,v)$ (or $Q_{n_1,n_2,k_n}(z;u,v)$) be the conditional probability that MK_n-NN rule classifies the observation Z into Π_1 (or Π_2), given $Z=z$, $U_n=u$, $V_n=v$. Then

$$(2.65)\quad P_{n_1,n_2,k_n}(z;u,v)$$

$$\underset{a.s.}{\doteq} (a_{n,1}+a_{n,2}+a_{n,3})/(a_{n,1}+a_{n,2}+a_{n,3}+b_{n,1}+b_{n,2}+b_{n,3})$$

$$Q_{n_1,n_2,k_n}(z;u,v) = 1-P_{n_1,n_2,k_n}(z;u,v).$$

As before, we let $\lambda = \lim n_2/n_1$ and $\theta = f_2(z)/f_1(z)$.

<u>Lemma 2.9.</u> Suppose that z is a continuous point of both f_1 and f_2 with $f_1(z) > 0$, $f_2(z) > 0$. If $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$, and $0 < \lambda < \infty$, then

$$(2.66)\quad \operatorname*{plim}_{n\to\infty} a_{n,1}/b_{n,1} = \begin{cases} 0 & \text{if } \lambda\theta > 1 \\ \infty & \text{if } \lambda\theta < 1 \end{cases}, \quad i=1,2,3.$$

<u>Proof.</u> We shall prove for $i = 1$, and $\lambda\theta < 1$ since others are of the same type and can be similarly proved. Let $\gamma_n = k_n-1$, and

$$(2.67)\quad \alpha(u,v) = \frac{F_1(u)-F_1(u)}{F_2(v)-F_2(u)} \cdot \frac{1-(F_2(v)-F_2(u))}{1-(F_1(v)-F_1(u))}$$

then

$$\frac{a_{n,1}}{b_{n,1}} = \frac{\sum_{j=\gamma_n}^{2\gamma_n+1} \binom{n_1-2}{j}\binom{n_2}{2\gamma_n+1-j}\alpha^j(u,v)}{\sum_{j=0}^{\gamma_n-1} \binom{n_1-2}{j}\binom{n_2}{2\gamma_n+1-j}\alpha^j(u,v)}$$

$$= \frac{\sum_{j=0}^{\gamma_n+1} \binom{n_1-2}{\gamma_n+j}\binom{n_2}{\gamma_n+1-j}\alpha^j(u,v)}{\sum_{j=1}^{\gamma_n} \binom{n_1-2}{\gamma_n-j}\binom{n_2}{\gamma_n+1+j}\alpha^{-j}(u,v)}$$

$$= \frac{\sum_{j=0}^{\gamma_n+1} \binom{2\gamma_n+1}{\gamma_n+j}\frac{(n_1-2-\gamma_n)!}{(n_1-2-\gamma_n-j)!}\frac{(n_2-\gamma_n-1)!}{(n_2-\gamma_n-1+j)!}\alpha^j(u,v)}{\sum_{j=1}^{\gamma_n} \binom{2\gamma_n+1}{\gamma_n-j}\frac{(n_1-2-\gamma_n)!}{(n_1-2-\gamma_n+j)!}\frac{(n_2-\gamma_n-1)!}{(n_2-\gamma_n-1-j)!}\alpha^{-j}(u,v)}$$

By Lemma 2.8, $\alpha(u,v) \xrightarrow{P} \theta^{-1}$ as $n\to\infty$ and $kn/n\to 0$ as $n\to\infty$. Suppose $\alpha(u,v)\to\theta^{-1}$ as $n\to\infty$. If $\lambda\theta < 1$, considering u,v as non-stochastic, there exists a constant $c(1 < c < 1/\lambda\theta)$ and a positive integer N such that for all $n \geq N$, we have

$$(2.68)\quad \frac{(n_1-2-\gamma_n)!}{(n_1-2-\gamma_n-j)!}\frac{(n_2-\gamma_n-1)!}{(n_2-\gamma_n-1+j)!}\alpha^j(u,v) \qquad j=0,1,\ldots,(\gamma_n+1)$$

$$\geq \{[(n_1-1-\gamma_n-j)/(n_2-\gamma_n-1+j)]\alpha(u,v)\}^j$$

$$\geq \{[(n_1-2-2\gamma_n)/n_2]\alpha(u,v)\}^j$$

$$\geq c^j$$

Similarly,

$$(2.69)\quad \frac{(n_1-2-\gamma_n)!}{(n_1-2-\gamma_n+j)!}\frac{(n_2-\gamma_n-1)!}{(n_2-\gamma_n-1-j)!}\alpha^{-j}(u,v) \quad j=1,\ldots,\gamma_n.$$

$$\leq \{[(n_1-2-\gamma_n+j)/(n_2-\gamma_n-j)]\,\alpha(u,v)\}^{-j}$$

$$\leq \{[(n_1-2)/(n_2-2\gamma_n)]\ \alpha(u,v)\}^{-j}$$

$$\leq c^{-j}$$

Hence for $n \geq N$, we have

$$a_{n,1}/b_{n,1} \geq \sum_{j=0}^{\gamma_n+1} \binom{2\gamma_n+1}{\gamma_n+j} c^j \Big/ \sum_{j=1}^{\gamma_n} \binom{2\gamma_n+1}{\gamma_n-j} c^{-j}$$

Moreover,

$$\sum_{j=0}^{\gamma_n+1} \binom{2\gamma_n+1}{\gamma_n+j} c^j = c^{-\gamma_n} \sum_{j=\gamma_n}^{2\gamma_n+1} \binom{2\gamma_n+1}{j} c^j$$

$$= c^{-\gamma_n}(1+c)^{2\gamma_n+1} \sum_{j=\gamma_n}^{2\gamma_n+1} \binom{2\gamma_n+1}{j} \left(\frac{c}{1+c}\right)^j \left(\frac{1}{1+c}\right)^{2\gamma_n+1-j}$$

$$\geq [(1+c)/\sqrt{c}]^{2\gamma_n}\ (1+c)/2,$$

since $c/1+c > \frac{1}{2}$ and

$$\sum_{j=1}^{\gamma_n} \binom{2\gamma_n+1}{\gamma_n-j} c^{-j} \leq \sum_{j=1}^{\gamma_n} \binom{2\gamma_n+1}{\gamma_n-j} \leq \tfrac{1}{2}(2)^{2\gamma_n+1} = 2^{2\gamma_n}$$

Therefore for $n \geq N$

$$(2.70)\quad a_{n,1}/b_{n,1} = [(1+c)/2\sqrt{c}]^{2\gamma_n}[(1+c)/2] \geq [(1+c)2\sqrt{c}]^{2\gamma_n}$$

The expression in the right hand side of (2.70) tends to ∞ as $n\to\infty$.

Hence when $\lambda\theta < 1$

$$\operatorname*{plim}_{n\to\infty} a_{n,1}/b_{n,1} = \infty.$$

The proof is complete.

Theorem 2.2. Under the assumptions of Lemma 2.9, we have

$$(2.71)\quad \lim_{n\to\infty} P_{n_1,n_2,k_n}(z) = \lim_{n\to\infty}(1-Q_{n_1,n_2,k_n}(z)) = \begin{cases} 1 & \text{if } \lambda\theta < 1 \\ 0 & \text{if } \lambda\theta > 1 \end{cases}$$

Proof. $\operatorname*{plim}_{n\to\infty} P_{n_1,n_2,k_n}(z;u,v)$

$$= \operatorname*{plim}_{n\to\infty} (a_{n,1}+a_{n,2}+a_{n,3})/(a_{n,1}+a_{n,2}+a_{n,3}+b_{n,1}+b_{n,2}+b_{n,3})$$

$$= \operatorname*{plim}_{n\to\infty} \sum_{i=1}^{3} [a_{n,i}/(a_{n,i}+b_{n,i})][(a_{n,i}+b_{n,i})/(a_{n,1}+a_{n,2}+a_{n,3}$$

$$+b_{n,1}+b_{n,2}+b_{n,3})]$$

From Lemma 2.9, we have

$$\operatorname*{plim}_{n\to\infty} a_{n,i}/(a_{n,i}+b_{n,i}) = \begin{cases} 1 & \text{if } \lambda\theta < 1 \\ 0 & \text{if } \lambda\theta > 1 \end{cases} \qquad i=1,2,3.$$

Therefore,

$$\operatorname*{plim}_{n\to\infty} P_{n_1,n_2,k_n}(z;u,v) = \begin{cases} 1 & \text{if } \lambda\theta < 1 \\ 0 & \text{if } \lambda\theta > 1. \end{cases}$$

Thus, by the Lebesque dominated convergence theorem, we get

$$\operatorname*{plim}_{n\to\infty} P_{n_1,n_2,k_n}(z) = \begin{cases} 1 & \text{if } \lambda\theta < 1 \\ 0 & \text{if } \lambda\theta > 1. \end{cases}$$

In order to apply the result we need to assume that $\lambda\theta \neq 1$ a.e. Furthermore, if the training samples are drawn from a population which is a mixture of π_1 and π_2 in the proportion ξ_1 and ξ_2, $(\lambda = \xi_2/\xi_1)$ then the asymptotic risk of the MK_n-NN rule is

$$\widetilde{R} = \int_{\{\xi_2 f_2 > \xi_1 f_1\}} \xi_1 f_1(z)dz + \int_{\{\xi_2 f_2 < \xi_1 f_1\}} \xi_2 f_2(z)dz$$

$$= \int \min\{\xi_1 f_1(z),\ \xi_2 f_2(z)\}dz = R^*, \text{ the Bayes risk.}$$

CHAPTER 3

ASYMPTOTIC PMC WITH RATE OF SOME SPECIFIC RULES BASED ON U-STATISTICS

3.0 Introduction

Consider a random variable X which is distributed as F_i in the population $\pi_i (i=0,1,2)$. The problem is to decide between $F_0=F_1$ and $F_0=F_2$ when it is known that F_1 and F_2 are different.

Let $\underset{\sim}{X}_{0n_0}=(X_{01}, \ldots, X_{0n_0})$ be n_0 independent observations on X from the population π_0, $\underset{\sim}{X}_{1n_1}=(X_{11}, \ldots, X_{1n_1})$ be n_1 independent observations on X from the population π_1, and $\underset{\sim}{X}_{2n_2}=(X_{21}, \ldots, X_{2n_2})$ be n_2 independent observations on X from the population π_2.

Define a function C as

$$(3.1) \quad C(u) = \begin{matrix} 1 & \text{if} & u > 0 \\ 0 & \text{if} & u \leq 0 \end{matrix}$$

The Wilcoxon statistics W_{01}, W_{02}, W_{12} are then defined as follows:

$$(3.2)\quad W_{01} = \frac{1}{n_0 n_1} \sum_{\substack{1\le i\le n_0 \\ 1\le j\le n_1}} c(X_{0i}-X_{1j})$$

$$W_{02} = \frac{1}{n_0 n_2} \sum_{\substack{1\le i\le n_0 \\ 1\le k\le n_2}} c(X_{0i}-X_{2k})$$

$$W_{12} = \frac{1}{n_1 n_2} \sum_{\substack{1\le j\le n_1 \\ 1\le k\le n_2}} c(X_{1j}-X_{2k})$$

Das Gupta (1964) considers a classification rule which decides $\pi_0=\pi_i$ $(i=1,2)$ if $|W_{0i}-\frac{1}{2}| = \min_{j=1,2} |W_{0j}-\frac{1}{2}|$. Under slight restriction on the distribution functions that $\int F_1 dF_2 > \frac{1}{2}$, Hudimoto (1964) also proposes a rule which is equivalent to classifying π_0 into π_1 if $(W_{01}+W_{02}-1) < 0$. (By symmetry, if $\int F_2 dF_1 > \frac{1}{2}$ is assumed, decide $\pi_0=\pi_2$ when $(W_{01}+W_{02}-1) < 0$). When it is not certain that whether $\int F_1 dF_2 > \frac{1}{2}$ or $\int F_2 dF_1 > \frac{1}{2}$, Chanda and Lee (1975), modifying Hudimoto's rule, suggest a rule which decides $\pi_0=\pi_1$ if $(W_{12}-\frac{1}{2})(W_{01}+W_{02}-1) > 0$.

In this chapter, the asymptotic probabilities of misclassification of the three rules mentioned above are obtained together with the rate of convergence when n_1 and n_2 approach infinity with n_0 fixed. Also Hudimoto's idea is applied to general classification problems. An example of a two-sided classification problem which utilizes the Lehmann statistic (Lehmann, 1951) is given. The asymptotic PMC and an upper bound of PMC are shown for this specific example as well.

3.1 Preliminaries.

Suppose we have a U-statistic (see Fraser (1957), pp. 223-224) defined by

$$(3.3)\quad U \equiv U(\underset{\sim}{X}_{0n_0};\ \underset{\sim}{X}_{1n_1};\ \underset{\sim}{X}_{2n_2})$$

$$= \frac{1}{\binom{n_0}{m_0}\binom{n_1}{m_1}\binom{n_2}{m_2}} \sum_{\substack{1\le\alpha_1<\ldots<\alpha_{m_0}\le n_0 \\ 1\le\beta_1<\ldots<\beta_{m_1}\le n_1 \\ 1\le\gamma_1<\ldots<\gamma_{m_2}\le n_2}} f(X_{0\alpha_1},\ldots,X_{0\alpha_{m_0}};\ X_{1\beta_1},\ldots,X_{1\beta_{m_1}};\ X_{2\gamma_1},\ldots,X_{2\gamma_{m_2}})$$

with $m_i \le n_i$, $i=0,1,2$.

Define

$$(3.4)\quad h(X_{1\beta_1},\ldots,X_{1\beta_{m_1}};X_{2\gamma_1},\ldots,X_{2\gamma_{m_2}}|\underset{\sim}{X}_{0n_0})$$

$$= \binom{n_0}{m_0}^{-1} \sum_{1\le\alpha_1<\ldots<\alpha_{m_0}\le n_0} f(X_{0\alpha_1},\ldots,X_{0\alpha_{m_0}};X_{1\beta_1},\ldots,X_{1\beta_{m_1}};\ X_{2\gamma_1},\ldots,X_{2\gamma_{m_2}})$$

Then U can be written as

$$(3.5)\quad U = \binom{n_1}{m_1}^{-1}\binom{n_2}{m_2}^{-1} \sum_{\substack{1\le\beta_1<\ldots<\beta_{m1}\le n_1 \\ 1\le\gamma_1<\ldots<\gamma_{m2}\le n_2}} h(X_{1\beta_1},\ldots,X_{1\beta_{m_1}};\ X_{2\gamma_1},\ldots,X_{2\gamma_{m_2}}|\underset{\sim}{X}_{0n_0})$$

which is also a U-statistic in $\underset{\sim}{X}_{1n_1}$ and $\underset{\sim}{X}_{2n_2}$ considering $\underset{\sim}{X}_{0n_0}$ as fixed. We define a function $h_{c_1,c_2}(x_{11},\ldots,x_{1c_1};x_{21},\ldots,x_{2c_2}|\underset{\sim}{X}_{0n_0})$ by taking the conditional expectation of $h(X_{11},\ldots,X_{1m_1};X_{21},\ldots,X_{2m_2}|\underset{\sim}{X}_{0n_0})$ given $x_{11},\ldots,x_{1c_1}$; $x_{21},\ldots,x_{2c_2}$:

$$(3.6)\quad h_{c_1,c_2}(x_{11},\ldots,x_{1c_1};x_{21},\ldots,x_{2c_2})$$
$$= \mathcal{E}\{h(x_{11},\ldots,x_{1c_1},X_{1c_1+1},\ldots,X_{1m_1};x_{21},\ldots,x_{2c_2},$$
$$X_{2c_2+1},\ldots,X_{2m_2}|\underset{\sim}{X}_{0n_0})\}$$

for $c_i=0,1,\ldots,m_i$; $i=1,2$.

In particular, define

$$(3.7)\quad \Psi(\underset{\sim}{X}_{0n_0}) = \mathcal{E}\{h(\quad;\quad|\underset{\sim}{X}_{0n_0})\}$$

Let

$$(3.8)\quad \zeta_{c_1,c_2}(\underset{\sim}{X}_{0n_0}) = \mathrm{var}[h_{c_1,c_2}(X_{11},\ldots,X_{1c_1};X_{21},\ldots,X_{2c_2}|\underset{\sim}{X}_{0n_0})]$$

and

$$(3.9)\quad \sigma^2(\underset{\sim}{X}_{0n_0}) = (m_1^2/n_1)\zeta_{1,0}(\underset{\sim}{X}_{0n_0})+(m_2^2/n_2)\zeta_{0,1}(\underset{\sim}{X}_{0n_0})$$

which can be expressed as

$$\sigma^2(\underset{\sim}{X}_{0n_0}) = \frac{1}{N}\{(m_1^2N/n_1)\zeta_{1,0}(\underset{\sim}{X}_{0n_0})+(m_2^2N/n_2)\zeta_{0,1}(\underset{\sim}{X}_{0n_0})\}$$

where

(3.10) $N = \min\{n_1,n_2\}$

Therefore if $\lim N/n_i$ exists, $\sigma^2(\underset{\sim}{X}_{0n_0})$ can asymptotically be written as

(3.11) $\sigma^2(\underset{\sim}{X}_{0n_0}) \doteq 1/M^2\varphi^2(\underset{\sim}{X}_{0n_0})$, $M^2 = N$

here $\varphi^2(\underset{\sim}{X}_{0n_0}) = \lim\limits_{\substack{n_1\to\infty \\ n_2\to\infty}} 1/\{(m_1^2N/n_1)\zeta_{1,0}(\underset{\sim}{X}_{0n_0})+(m_2^2N/n_2)\zeta_{0,1}(\underset{\sim}{X}_{0n_0})\}$

is a function of $\underset{\sim}{X}_{0n_0}$ not depending on N.

Using the notation introduced above, we now give the following proposition.

<u>Proposition 3.1</u> If $\lim N/n_i$ exists and assume that with probability one

(3.12) $\mathcal{E}|h(\quad ; \quad |\underset{\sim}{X}_{0n_0})|^2 < \infty$

and

(3.13) $N\sigma^2(\underset{\sim}{X}_{0n_0}) > 0$ as $N \to \infty$

Then

(3.14) $\Pr\{U<0\} \to \Pr\{\Psi(\underset{\sim}{X}_{0n_0})<0\}+\frac{1}{2}\Pr(\Psi(\underset{\sim}{X}_{0n_0})=0)$ as $N \to \infty$

Proof. Conditioning on $\underset{\sim}{X}_{0n_0}$, asymptotic normality theorems (Hoeffding (1948), Lehmann (1951)) for U-statistics state that

(3.15) $\Pr\{U<0|\underset{\sim}{X}_{0n_0}\} \doteq \Phi(-\Psi(\underset{\sim}{X}_{0n_0})/\sigma(\underset{\sim}{X}_{0n_0}))$ as $N \to \infty$

Combining with (3.11), we have

(3.16) $\Pr\{U<0|\underset{\sim}{X}_{0n_0}\} \doteq \Phi(-M\Psi(\underset{\sim}{X}_{0n_0})\varphi(\underset{\sim}{X}_{0n_0}))$ as $N \to \infty$

Hence

(3.17) $\Pr\{U<0\} = \mathcal{E}\Pr\{U<0|\underset{\sim}{X}_{0n_0}\}$

$\to \Pr\{\Psi(\underset{\sim}{X}_{0n_0})\varphi(\underset{\sim}{X}_{0n_0})<0\}+\frac{1}{2}\Pr\{\Psi(\underset{\sim}{X}_{0n_0})\varphi(\underset{\sim}{X}_{0n_0}) = 0\}$ as $N \to \infty$

$=\Pr\{\Psi(\underset{\sim}{X}_{0n_0})<0\}+\frac{1}{2}\Pr\{\Psi(\underset{\sim}{X}_{0n_0}) = 0\}$

since $\varphi(\underset{\sim}{X}_{0n_0})$ is positive with probability one.

Proposition 3.2 If $\lim N/n_i$ exists and assume that

(3.18) $\mathcal{E}|f|^3 < \infty$,

(3.19) $\mathcal{E}|f|^{2r} < \infty$ for a positive r

(3.20) for sufficiently small $\epsilon > 0$, $\Pr\{|\Psi(\underset{\sim}{X}_{On_O})| \leq \epsilon\} = O(\epsilon)$

and

(3.21) $N\sigma^2(\underset{\sim}{X}_{On_O}) \geq a > 0$ as $N \to \infty$

Then

(3.22) $\Pr\{U<0\} = \Pr\{\Psi(\underset{\sim}{X}_{On_O})<0\}+O(N^{-r/(2r+1)})$ as $N \to \infty$

__Proof.__ Conditioning on $\underset{\sim}{X}_{On_O}$ and following the proof of Theorem 3.1 of Grams and Serfling (1973) with (3.21) we have

(3.23) $\Pr\{U<0|\underset{\sim}{X}_{On_O}\} = \Phi(-\Psi(\underset{\sim}{X}_{On_O})/\sigma(\underset{\sim}{X}_{On_O}))+N^{-r/(2r+1)}K(\underset{\sim}{X}_{On_O})$

as $N \to \infty$

$= \Phi(-M\Psi(\underset{\sim}{X}_{On_O})\varphi(\underset{\sim}{X}_{On_O}))+N^{-r/(2r+1)}K(\underset{\sim}{X}_{On_O})$ as $N \to \infty$

where $K(\underset{\sim}{X}_{On_O})$ is a function (independent of N) depending on $\underset{\sim}{X}_{On_O}$ through the 3rd and 2rth absolute moments of $(h_{O,1}(\quad ; \quad |\underset{\sim}{X}_{On_O})-\Psi(\underset{\sim}{X}_{On_O}))$ and $(h_{1,O}(\quad ; \quad |\underset{\sim}{X}_{On_O}) -\Psi(\underset{\sim}{X}_{On_O}))$.

Hence (3.18), (3.19), and (3.23) imply

$$(3.24)\quad \Pr\{U<0\} = \mathcal{E}\Phi(-M\Psi(\underset{\sim}{X}_{0n_0})\varphi(\underset{\sim}{X}_{0n_0}))+O(N^{-r/(2r+1)}) \quad \text{as} \quad N\to\infty$$

Since $\Phi(-M) = O(M^{-k})$ for any positive k as $M\to\infty$ we have for every $\epsilon > 0$, as $N\to\infty$,

$$(3.25)\quad \Pr\{U<0\} = \int_{-\infty}^{-M^{-1+\epsilon}} \Phi(-M\Psi\varphi)dp(\Psi\varphi)+\int_{-M^{-1+\epsilon}}^{M^{-1+\epsilon}}\Phi(-M\Psi\varphi)dp(\Psi\varphi)$$

$$+\int_{M^{-1+\epsilon}}^{\infty} \Phi(-M\Psi\varphi)dp(\Psi\varphi)+O(N^{-r/(2r+1)})$$

$$= \Pr\{\Psi\varphi \le -M^{-1+\epsilon}\}+O(N^{-r/(2r+1)})$$

$$= \Pr\{\Psi\varphi<0\}-\Pr\{-M^{-1+\epsilon} < \Psi\varphi < 0\}+O(N^{-r/(2r+1)})$$

$$= \Pr\{\Psi(\underset{\sim}{X}_{0n_0})<0\}+O((M^{-1+\epsilon}))+O(N^{-r/(2r+1)})$$

$$= \Pr\{\Psi(\underset{\sim}{X}_{0n_0})<0\}+O(N^{-r/(2r+1)})$$

because ϵ is arbitrary.

Corollary 3.2 Assume that f has finite moments of all orders. If (3.20) and (3.21) hold, then for every $\epsilon > 0$

$$(3.26)\quad \Pr\{U<0\} = \Pr\{\Psi(\underset{\sim}{X}_{0n_0})<0)+O(N^{-\frac{1}{2}+\epsilon}) \quad \text{as} \quad N\to\infty$$

3.2 Asymptotic PMC's

Let $P_N(D)$, $P_N(H)$, $P_N(C)$ be the probabilities of classifying π_0 into π_1 for Das Gupta's rule, Hudimoto's rule, and Chanda and Lee's rule, respectively. Note that $1-P_N(\cdot)$ is the PMC if $\underset{\sim}{X}_{0n_0} \sim F_1$ and $P_N(\cdot)$ is the PMC if $\underset{\sim}{X}_{0n_0} \sim F_2$. Therefore, to study PMC, it is sufficient to study $P_N(\cdot)$.

If the conditions (3.20), (3.21) are assumed to be satisfied, and since the functions are all bounded (in fact between -1 and 1) the moments of all orders are finite. Moreover, the product of U-statistics is again a U-statistic. Then from Corollary 3.2 we have, for every $\epsilon > 0$,

$$(3.29)\quad P_N(D) = \Pr\{|W_{01}-\tfrac{1}{2}|<|W_{02}-\tfrac{1}{2}|\}$$

$$= \Pr\{(W_{01}-W_{02})(W_{01}+W_{02}-1)<0\}$$

$$= \Pr\{[\frac{1}{n_0}\sum_{i=1}^{n_0}(F_1(X_{0i})-F_2(X_{0i}))][\frac{1}{n_0}\sum_{i=1}^{n_0}(F_1(X_{0i})+F_2(X_{0i})-1)]<0\} + O(N^{-\frac{1}{2}+\epsilon}) \quad \text{as } N\to\infty$$

$$(3.30)\quad P_N(H) = \Pr\{W_{01}+W_{02}-1 < 0\}$$

$$= \Pr\{\frac{1}{n_0}\sum_{i=1}^{n_0}(F_1(X_{0i})-F_2(X_{0i}))-1<0\} + O(N^{-\frac{1}{2}+\epsilon}) \text{ as } N\to\infty$$

$$(3.31)\quad P_N(C) = \Pr\{(W_{12}-\tfrac{1}{2})(W_{01}+W_{02}-1) >0\}$$

$$= \begin{cases} \Pr\{\frac{1}{n_0}\sum_{i=1}^{n_0}(F_1(X_{0i})+F_2(X_{0i}))-1<0\}+O(N^{-\frac{1}{2}+\epsilon}) \\ \quad \text{if } \int F_1 dF_2 > \frac{1}{2} \\ \Pr\{\frac{1}{n_0}\sum_{i=1}^{n_0}(F_1(X_{0i})+F_2(X_{0i})-1>0\}+O(N^{-\frac{1}{2}+\epsilon}) \\ \quad \text{if } \int F_1 dF_2 < \frac{1}{2} \end{cases}$$

as $N \to \infty$

3.3 A general rule based on U-statistics and an example.

We shall generally describe Hudimoto's idea. Suppose we have U-statistics V_1 and V_2 defined by

$$(3.32)\quad V_1 \equiv V_1(\underset{\sim}{X}_{0n_0};X_{1n_1}) = \frac{1}{\binom{n_0}{m_0}\binom{n_1}{m_1}} \sum_{\substack{1\le\alpha_1<\dots<\alpha_{m_0}\le n_0 \\ 1\le\beta_1<\dots<\beta_{m_1}\le n_1}} f_1(X_{0\alpha_1},\dots,X_{0\alpha_{m_0}};X_{1\beta_1},\dots,X_{1\beta_{m_1}})$$

$$(3.33)\quad V_2 \equiv V_2(\underset{\sim}{X}_{0n_0};X_{2n_2}) = \frac{1}{\binom{n_0}{m_0}\binom{n_2}{m_2}} \sum_{\substack{1\le\alpha_1<\dots<\alpha_{m_0}\le n_0 \\ 1\le\gamma_1<\dots<\gamma_{m_2}\le n_2}} f_1(X_{0\alpha_1},\dots,X_{0\alpha_{m_0}};X_{2\gamma_1},\dots,X_{2\gamma_{m_2}})$$

such that for some $\theta_1 < \theta_2$ we have

$$(3.34)\quad \mathcal{E}(V_1|\underset{\sim}{X}_{0n_0} \in \Pi_1) = \theta_1, \quad \mathcal{E}(V_2|\underset{\sim}{X}_{0n_0} \in \Pi_1) = \theta_2$$

and

$$(3.35)\quad \mathcal{E}(V_1|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \theta_2, \quad \mathcal{E}(V_2|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \theta_1$$

Then the rule will be the one that classifies Π_0 as Π_1 if $V_1 < V_2$ and Π_2 if $V_1 \geq V_2$.

Let

$$(3.36)\quad \theta = \theta_2 - \theta_1, \qquad n = \min\{n_0, n_1, n_2\},$$

$$(3.37)\quad f(X_{01},\ldots,X_{0m_0};X_{11},\ldots,X_{1m_1};X_{21},\ldots,X_{2m_2})$$

$$= f_1(X_{01},\ldots,X_{0m_0};X_{11},\ldots,X_{1m_1}) - f_1(X_{01},\ldots,X_{0m_0};$$

$$X_{21},\ldots,X_{2m_2}),$$

and

$$(3.38)\quad U_n \equiv U_n(\underset{\sim}{X}_{0n_0};\underset{\sim}{X}_{1n_1};\underset{\sim}{X}_{2n_2}) = V_1 - V_2$$

then

$$(3.39)\quad U_n = \frac{1}{\binom{n_0}{m_0}\binom{n_1}{m_1}\binom{n_2}{m_2}} \sum_{\substack{1\leq\alpha_1<\ldots<\alpha_{m_0}\leq n_0 \\ 1\leq\beta_1<\ldots<\beta_{m_1}\leq n_1 \\ 1\leq\gamma_1<\ldots<\gamma_{m_2}\leq n_2}} f(X_{0\alpha_1},\ldots,X_{0\alpha_{m_0}}; X_{1\beta_1},\ldots,X_{1\beta_{m_1}}; X_{2\gamma_1},\ldots,X_{2\gamma_{m_2}})$$

and

$$(3.40)\quad \mathcal{E}(U_n|\underset{\sim}{X}_{0n_0} \in \Pi_1) = -\theta, \quad \mathcal{E}(U_n|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \theta$$

Consequently, the rule is simply classifying Π_0 into Π_1 if $U_n < 0$ and Π_2 if $U_n \geq 0$. Also U_n is again a U-statistic.

Regarding the asymptotic probabilities of misclassification as the sizes of the training samples tend to infinity, we shall refer to Proposition 3.1, Proposition 3.2, or Corollary 3.2.

We now give applications of the above results to a specific example based on the Lehmann statistic (Lehmann (1951)). This example is constructed for general two-sided classification problems. Only continuity and distinctness of the distribution functions are assumed.

We define the measure of discrepancy between two distribution functions F_1 and F_2 as

$$(3.41)\quad \Delta(F_1,F_2) = \int (F_1-F_2)^2 \, d\,\frac{F_1+F_2}{2}$$

Lehmann (1951) proves the following:

<u>Lemma 3.1</u> $F_1 = F_2$ iff $\Delta(F_1,F_2) = 0$

Let X_1, X_2 be independent random variables with distribution function F_1, and let Y_1, Y_2 be independent random variables with distribution function F_2. We designate $\max(X_1,X_2)$ as $X_1 \vee X_2$, and $\min(X_1,X_2)$ as $X_1 \wedge X_2$. When (X_1,X_2) and

(Y_1,Y_2) are independent, Lehmann (1951) proves:

<u>Lemma 3.2</u>

(3.42) $\Pr\{X_1 \vee X_2 < Y_1 \wedge Y_2 \text{ or } Y_1 \vee Y_2 < X_1 \wedge X_2\} = 1/3 + 2\Delta(F_1,F_2)$

From (3.42) we see immediately that

(3.43) $0 \leq \Delta(F_1,F_2) \leq 1/3$

Consider the statistics V_1 and V_2

(3.44) $$V_1 = \frac{1}{\binom{n_0}{2}\binom{n_1}{2}} \sum_{\substack{1\leq\alpha_1<\alpha_2\leq n_0 \\ 1\leq\beta_1<\beta_2\leq n_1}} g(X_{0\alpha_1},X_{0\alpha_2};X_{1\beta_1},X_{1\beta_2})$$

(3.45) $$V_2 = \frac{1}{\binom{n_0}{2}\binom{n_2}{2}} \sum_{\substack{1\leq\alpha_1<\alpha_2\leq n_0 \\ 1\leq\gamma_1<\gamma_2\leq n_2}} g(X_{0\alpha_1},X_{0\alpha_2};X_{2\gamma_1},X_{2\gamma_2})$$

where

(3.46) $$g(X_1,X_2;Y_1,Y_2) = \begin{cases} 1 & \text{if } X_1 \vee X_2 < Y_1 \wedge Y_2 \\ & \text{or } X_1 \wedge X_2 > Y_1 \vee Y_2 \\ 0 & \text{otherwise} \end{cases}$$

Define

$$(3.47)\quad U_n = V_1 - V_2$$

$$= \frac{1}{\binom{n_0}{2}\binom{n_1}{2}\binom{n_2}{2}} \sum_{\substack{1\le\alpha_1<\alpha_2\le n_0 \\ 1\le\beta_1<\beta_2\le n_1 \\ 1\le\gamma_1<\gamma_2\le n_2}} (g(X_{0\alpha_1},X_{0\alpha_2};X_{1\beta_1},X_{1\beta_2}) - g(X_{0\alpha_1},X_{0\alpha_2};X_{2\gamma_1},X_{2\gamma_2}))$$

$$\equiv \frac{1}{\binom{n_0}{2}\binom{n_1}{2}\binom{n_2}{2}} \sum_{\substack{1\le\alpha_1<\alpha_2\le n_0 \\ 1\le\beta_1<\beta_2\le n_1 \\ 1\le\gamma_1<\gamma_2\le n_2}} f(X_{0\alpha_1},X_{0\alpha_2};X_{1\beta_1},X_{1\beta_2};X_{2\gamma_1},X_{2\gamma_2})$$

Then from Lemma 3.1 and Lemma 3.2 we have

$$(3.48)\quad \mathcal{E}(U_n|\underset{\sim}{X}_{0n_0} \in \pi_1) = -2\Delta, \quad \mathcal{E}(U_n|\underset{\sim}{X}_{0n_0} \in \pi_2) = 2\Delta$$

Therefore, the rule is to classify π_0 into π_1 or π_2 according to $U_n < 0$ or $U_n \ge 0$. Note that $-1 \le f \le 1$. And for given X_{01}, X_{02}, we have

$$(3.49)\quad \mathcal{E}[f(X_{01},X_{02};X_{11},X_{12};X_{21},X_{22})|X_{01},X_{02}]$$

$$= \Pr\{X_{01}\vee X_{02} < X_{11}\wedge X_{12} \text{ or } X_{01}\wedge X_{02} > X_{11}\vee X_{12}|X_{01},X_{02}\}$$

$$- \Pr\{X_{01}\vee X_{02} < X_{21}\wedge X_{22} \text{ or } X_{01}\wedge X_{02} > X_{21}\vee X_{22}|X_{01},X_{02}\}$$

$$= [1-F_1(X_{01}\vee X_{02})]^2+F_1^2(X_{01}\wedge X_{02})-[1-F_2(X_{01}\vee X_{02})]^2- F_2^2(X_{01}\wedge X_{02})$$

Thus P_N, the probability of classifying Π_0 into Π_1, is found from Corollary 3.2. to be

$$(3.50)\ P_N = \Pr\{\frac{1}{\binom{n_0}{2}} \sum_{1\leq\alpha_1<\alpha_2\leq n_0} [(1-F_1(X_{0\alpha_1} \vee X_{0\alpha_2}))^2+F_1^2(X_{0\alpha_1} \wedge X_{0\alpha_2})- (1-F_2(X_{0\alpha_1} \vee X_{0\alpha_2}))^2-F_2^2(X_{0\alpha_1} \wedge X_{0\alpha_2})] < 0\} + O(N^{-\frac{1}{2}+\epsilon})$$

for every $\epsilon > 0$ as $N \to \infty$

3.4 Closing Remark

Sen (1960), Hoeffding (1961), and Berk (1966) have the following lemma on the convergence of U-statistics.

Lemma 3.3 As $n \to \infty$, U_n converges almost surely to the parameter it estimates unbiasedly.

The strong consistency of the rules mentioned above is then an immediate consequence.

Utilizing Hoeffding's inequality for bounded U-statistics (Hoeffding (1963), p. 25), we obtain an upper bound of PMC of the rule, which is based on the Lehmann statistic.

$$(3.51)\ PMC_j \leq e^{-2[n/2]\Delta^2} \qquad j=1, 2$$

where the subscript j indicates that the probability is calculated under the assumption that $\Pi_0 = \Pi_j$, and $[x]$ denotes the largest integer less than or equal to x.

CHAPTER 4

SEQUENTIAL RULES

4.0 Introduction

In the first part of this chapter we shall consider sequential classification rules based on U-statistics with bounded kernels so that the sampling will terminate with probability one and the PMC's can be made smaller than any preassigned arbitrary positive constant. We have extracted the basic idea from the work by Hoeffding and Wolfowitz (1958) on distinguishability of sets of distributions. Later the notion of distinguishability was used by Das Gupta and Kinderman (1974) in the set-up for the classification problems. Hoeffding and Wolfowitz (1958) introduced the minimum distance test procedure and studied the properties of this test using the available probability bounds on sample distance function. We shall introduce the minimum-U sequential rules and prove some properties of these rules by using the available probability inequality for U-statistics.

In the second part of this chapter we shall consider some sequential rules when $F_1 = N_p(\mu_1, \Sigma)$ and $F_2 = N_p(\mu_2, \Sigma)$. Following the idea of Chow and Robbins (1965) and Simons (1968), Srivastava (1973) proposed some sequential rules for the following two cases: (i) $\mu_1 - \mu_2 = \delta$ known but Σ is unknown. (ii) Both δ and Σ are unknown. For the case (i) Srivastava (1973) proposed a sequential rule based on observations from π_0 and π_1 and, given α he showed that the PMC's of this rule tend to values less than α as $\delta'\Sigma^{-1}\delta \to 0$. Furthermore he modified his rule so that the prescribed errors can actually be achieved. He also proved the asymptotic efficiency of his rule. However Srivastava's proof is incomplete and suffers from a technical error. We shall present a more rigorous analysis of his rule (not the modified one). Srivastava also proved that for his rule in case (ii) the errors can be controlled arbitrarily as $\delta'\Sigma^{-1}\delta \to 0$. However his proof is entirely wrong and here we shall indicate his error. Unfortunately such an optimal result is not true in this case.

4.1 Minimum-U sequential rules.

As in Chapter 3, we want to decide between $F_0=F_1$ and $F_0=F_2$, and samples $\underset{\sim}{X}_{0n_0}$, $\underset{\sim}{X}_{1n_1}$, $\underset{\sim}{X}_{2n_2}$ are available from the populations π_0, π_1, π_2, respectively.

We shall study the problem in the following set-up. Consider U-statistics defined by

$$(4.1)\quad \begin{aligned} V_1 &\equiv V_1(\underset{\sim}{X}_{0n_0};\underset{\sim}{X}_{1n_1}) = \binom{n_0}{m}^{-1}\binom{n_1}{m}^{-1}\sum h(X_{0\alpha_1},\ldots,X_{0\alpha_m};\\ &\qquad\qquad X_{1\beta_1},\ldots,X_{1\beta_m})\\ V_2 &\equiv V_2(\underset{\sim}{X}_{0n_0};\underset{\sim}{X}_{2n_2}) = \binom{n_0}{m}^{-1}\binom{n_1}{m}^{-1}\sum h(X_{0\alpha_1},\ldots,X_{0\alpha_m};\\ &\qquad\qquad X_{2\gamma_1},\ldots,X_{2\gamma_m})\end{aligned}$$

for $m \leq n$; $n = \min\{n_0, n_1, n_2\}$, where the summation is over all possible combinations. Furthermore, assume that

$$(4.2)\quad \begin{array}{ll} \mathcal{E}(V_1|\underset{\sim}{X}_{0n_0} \in \Pi_1) = \theta_1, & \mathcal{E}(V_2|\underset{\sim}{X}_{0n_0} \in \Pi_1) = \theta_1+\theta \\ \mathcal{E}(V_1|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \theta_1+\theta, & \mathcal{E}(V_2|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \theta_1, \end{array}$$

where $\theta > 0$. It is also assume that h is bounded; namely, there exist d_1 and $d_2 (-\infty < d_1 < d_2 < \infty)$ such that $d_1 \leq h \leq d_2$. Then $d_1 \leq V_1, V_2 \leq d_2$.

To illustrate the above set-up, consider (Hudimoto, 1964)

$$(4.3)\quad \begin{array}{l} \hat{P}_1 = (1/n_0 n_1) \sum c(X_{0i}-X_{1j}) \\ \hat{P}_2 = (1/n_0 n_2) \sum c(X_{2k}-X_{0i}) \end{array}$$

where $c(x) = \begin{array}{ll} 1 & x > 0 \\ 0 & x \leq 0 \end{array}$

Hudimoto (1964) showed that

$$(4.4)\quad \begin{array}{ll} \mathcal{E}(\hat{P}_1|\underset{\sim}{X}_{0n_0} \in \Pi_1) = \tfrac{1}{2}, & \mathcal{E}(\hat{P}_2|\underset{\sim}{X}_{0n_0} \in \Pi_1) = \tfrac{1}{2}+\bar{\Delta} \\ \mathcal{E}(\hat{P}_1|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \tfrac{1}{2}+\bar{\Delta}, & \mathcal{E}(\hat{P}_2|\underset{\sim}{X}_{0n_0} \in \Pi_2) = \tfrac{1}{2}, \end{array}$$

where $\bar{\Delta} = \int F_1 dF_2 - \tfrac{1}{2} > 0$ (assuming F_1 and F_2 are distinct and continuous).

For another illustration, consider

$$(4.5)\quad \tilde{P}_1 = \binom{n_0}{2}^{-1} \binom{n_1}{2}^{-1} \sum g(X_{0\alpha_1}, X_{0\alpha_2}; X_{1\beta_1}, X_{1\beta_2})$$

$$\tilde{P}_2 = \binom{n_0}{2}^{-1}\binom{n_2}{2}^{-1} \sum g(X_{0\alpha_1}, X_{0\alpha_2}; X_{2\gamma_1}, X_{2\gamma_2})$$

where g is defined in (3.46). Then from Lemma 3.1 and Lemma 3.2, we get

$$(4.6)\quad \begin{aligned} &\mathcal{E}(\tilde{P}_1 | \underset{\sim}{X}_{0n_0} \epsilon \pi_1) = 1/3, &&\mathcal{E}(\tilde{P}_2 | \underset{\sim}{X}_{0n_0} \epsilon \pi_1) = 1/3+2\Delta \\ &\mathcal{E}(\tilde{P}_1 | \underset{\sim}{X}_{0n_0} \epsilon \pi_2) = 1/3+2\Delta, &&\mathcal{E}(\tilde{P}_2 | \underset{\sim}{X}_{0n_0} \epsilon \pi_2) = 1/3 \end{aligned}$$

where $\Delta = \int (F_1 - F_2)^2 d\frac{F_1+F_2}{2}$.

As before, we write U_n as

$$\begin{aligned} (4.7)\quad U_n &\equiv V_1 - V_2 \\ &= \binom{n_0}{m}^{-1}\binom{n_1}{m}^{-1}\binom{n_2}{m}^{-1} \sum f(X_{0\alpha_1}, \ldots, X_{0\alpha_m}; X_{1\beta_1}, \ldots, X_{1\beta_m}; \\ &\qquad\qquad X_{2\gamma_1}, \ldots, X_{2\gamma_m}), \end{aligned}$$

where $f(X_{01}, \ldots, X_{0m}; X_{11}, \ldots, X_{1m}; X_{21}, \ldots, X_{2m}) = h(X_{01}, \ldots, X_{0m};$
$X_{11}, \ldots, X_{1m}) - h(X_{01}, \ldots, X_{0m}; X_{21}, \ldots, X_{2m})$

Note that $-a \leq f \leq a$; $a = d_2 - d_1 > 0$.

In Chapter 3, we have considered the rule which classifies π_0 into π_1 if $U_n < 0$ and into π_2 if $U_n \geq 0$. From Hoeffding's inequality (1963, p. 25) for bounded U-statistics, we obtain an upper bound of PMC, which is given by

(4.8) $\exp(-[n/m]\theta^2/2a^2), \qquad \theta > 0.$

We shall consider two cases: (i) θ_1 is known but θ is unknown. (ii) Both θ_1 and θ are unknown.

4.1.1 <u>Minimum-U sequential rule I:</u> θ_1 <u>known.</u>

Often θ_1 in (4.2) is a known constant as we have seen in (4.4) and (4.6) where $\theta_1 = \frac{1}{2}$ and $\theta_1 = 1/3$, respectively. Without loss of generality we shall assume $\theta_1 = 0$. Then we define a sequential rule as follows:

First we choose a sequence $\{\alpha_n\}$ of positive constants such that

(4.9) $\sum_{i=1}^{\infty} \alpha_i \leq p$

and a sequence $\{C_i\}$ of positive numbers such that

(4.10) $\lim_{n\to\infty} C_n = 0, \quad \text{and} \quad 0 < C_n < d_2 \quad \text{for all} \quad n \geq 1.$

and a strictly increasing sequence $\{m_i\}$ of positive integers such that

(4.11) $\exp(-[m_i/m]C_i^2/2a^2) \leq \alpha_i$ for all $i \geq 1$ and $m_1 \geq m$.

Put

(4.12) $\delta_i = \max\{V_1(\underset{\sim}{X}_{0m_i};\underset{\sim}{X}_{1m_i}), \quad V_2(\underset{\sim}{X}_{0m_i};\underset{\sim}{X}_{2m_i})\}$

Take successive independent samples of sizes m_1, m_2-m_1,

$m_3 - m_2$, ... Continue sampling as long as $\delta_i < C_i$. Stop sampling as soon as $\delta_i \geq C_i$, and apply the terminal rule

$$(4.13)\quad \varphi = \begin{cases} 1 & \text{if } V_1(\underset{\sim}{X}_{0m_i};\underset{\sim}{X}_{1m_i}) < V_2(\underset{\sim}{X}_{0m_i};\underset{\sim}{X}_{2m_i}) \\ 0 & \text{otherwise} \end{cases}$$

We classify Π_0 into Π_1 or Π_2 according to $\varphi = 1$ or 0. Hence the sample size is

$$(4.14)\quad N = m_t,$$

where t is the first integer i for which $\delta_i \geq C_i$.

We shall denote this rule by (N, φ). Following the argument of Hoeffding and Wolfowitz (1958), we get the following results.

<u>Proposition 4.1.</u> The rule (N, φ) terminates with probability one.

<u>Proof.</u> It suffices to show that $\Pr\{N < \infty | \Pi_0 = \Pi_1\} = 1$ since $\Pr\{N < \infty | \Pi_0 = \Pi_2\}$ can be similarly proved.

$$\begin{aligned}(4.15)\quad \Pr\{N > m_j | \Pi_0 = \Pi_1\} &= \Pr\{\delta_i < C_i \quad \text{for} \quad 1 \leq i \leq j | \Pi_0 = \Pi_1\} \\ &\leq \Pr\{\delta_j < C_j | \Pi_0 = \Pi_1\} \\ &\leq \Pr\{V_2(\underset{\sim}{X}_{0m_j};\underset{\sim}{X}_{2m_j}) < C_j | \Pi_0 = \Pi_1\} \\ &= \Pr\{-V_2 + \mathcal{E}(V_2|\Pi_0 = \Pi_1) > \mathcal{E}(V_2|\Pi_0 = \Pi_1) - C_j | \Pi_0 = \Pi_1\} \\ &= \Pr\{-V_2 + \theta > \theta - C_j | \Pi_0 = \Pi_1\}\end{aligned}$$

Since $C_j \to 0$ and $\theta > 0$ we have $\theta - C_j > C_j$ for j sufficiently large, and then, by Hoeffding's inequality for U-statistics, the right side of (4.15) is $\leq \exp(-[m_j/m]C_j^2/2a^2) \leq \alpha_j$ (see (4.11)). By (4.9), $\alpha_j \to 0$ as $j \to \infty$. Thus $\Pr\{N > m_j\} \to 0$ as $j \to \infty$, which completes the proof.

<u>Proposition 4.2.</u> Each of the PMC's of the rule (N, φ) is less than p.

<u>Proof.</u> Since (N, φ) terminates we can write

$$\begin{aligned}
PMC_2 &= \mathcal{E}(\varphi | \Pi_0 = \Pi_2) \\
&= \sum_{i=1}^{\infty} \Pr\{\delta_j < C_j \text{ for } j < i,\ \delta_i \geq C_i,\ V_1 \leq V_2 | \Pi_0 = \Pi_2\} \\
&\leq \sum \Pr\{V_2 \geq C_i | \Pi_0 = \Pi_2\} \\
&\leq \sum \exp(-[m_i/m]C_i^2/2a^2) \qquad \text{by Hoeffding's inequality} \\
&\leq \sum_{i=1} \alpha_i \leq p
\end{aligned}$$

Similarly,

$$PMC_1 = \mathcal{E}(1 - \varphi | \Pi_0 = \Pi_1) \leq p.$$

Furthermore, if we choose the sequence $\{m_i\}$ suitably, the moment generating function of N will exist.

<u>Proposition 4.3.</u> If the sequence $\{m_i\}$ is so chosen that (in addition to (4.11))

(4.16) $\liminf_{i\to\infty} i^{-1}(2m[m_i/m]-m_{i+1}) > 0,$

then for every $\theta > 0$, there is a positive constant $t(\theta)$ such that $\mathcal{E}(\exp(tN)|\pi_0=\pi_i) < \infty$ for $t \le t(\theta)$, $i=1,2$.

Proof. We shall only prove that $\mathcal{E}(\exp(tN)|\pi_0=\pi_1) < \infty$. Since $C_j \to 0$ as $j \to \infty$ and $\theta > 0$, there exists a positive integer J such that $\theta - C_j > \theta/2$ for $j \ge J$. Therefore, for all $j \ge J$, due to (4.15) and Hoeffding's inequality, we have

(4.17) $\Pr\{N > m_j|\pi_0=\pi_1\} \le \exp(-[m_j/m](\theta/2)^2/2a^2)$

$$= \exp(-[m_j/m]\theta^2/8a^2)$$

Now for any real t,

$$\mathcal{E}(\exp(tN)|\pi_0=\pi_1) = \sum_{j=1}^{\infty} \exp(tm_j)\Pr\{N=m_j|\pi_0=\pi_1\}$$

$$\le \exp(tm_1) + \sum_{j=1}^{\infty} \exp(tm_{j+1}) \Pr\{N > m_j|\pi_0=\pi_1\}$$

Thus, from (4.17), $\mathcal{E}(\exp(tN)|\pi_0=\pi_1) < \infty$ if the series $\sum_{j=1}^{\infty} \exp(tm_{j+1})\exp(-[m_j/m]\theta^2/8a^2)$ converges. Since $\theta > 0$ and m is a positive integer, let $t(\theta) = \theta^2/16ma^2 > 0$. If $t \le t(\theta)$, then

$$tm_{j+1} - [m_j/m]\theta^2/8a^2 \le -(\theta^2/16ma^2)(2[m_j/m]-m_{j+1})$$

so that the series $\sum_{j=1}^{\infty} \exp(tm_{j+1})\exp(-[m_j/m]\theta^2/8a^2)$ converges due to (4.16). The proof is complete.

<u>Remark.</u> If, for a given $p > 0$, we choose $\alpha_j = p/2^j$, $C_j = dj^{-\frac{1}{2}}$, here $0 < d < d_2$, then $m_j \doteq 2m(a^2/d^2)(j^2\log 2+j\log(1/p))$. Therefore the conditions (4.11) and (4.16) hold, so the moment generating function of N exists.

4.1.2 <u>Minimum-U sequential rule II: θ_1 unknown.</u>

Define the sequences $\{\alpha_i\}$, $\{C_i\}$, $\{m_i\}$ as before and put

$$(4.18)\quad \Delta_i = |V_1(\underset{\sim}{X}_{0m_i};\underset{\sim}{X}_{1m_i})-V_2(\underset{\sim}{X}_{0m_i};\underset{\sim}{X}_{2m_i})|$$

Take samples of sizes m_1, m_2-m_1, m_3-m_2, ..., where m_i's are defined as in (4.11) Continue sampling as long as $\Delta_i < C_i$. Stop sampling as soon as $\Delta_i \geq C_i$ and apply the terminal rule φ (see (4.13)). The sample size is given by

$$(4.19)\quad N' = m_t,$$

where t is the first integer for which $\Delta_i \geq C_i$. We denote this rule by (N', φ).

<u>Proposition 4.1'</u> The rule (N', φ) terminates with probability one.

<u>Proof.</u> We shall only show $\Pr\{N < \infty | \Pi_0 = \Pi_1\} = 1$.

$$(4.20)\quad \Pr\{N > m_j|\pi_0=\pi_1\} = \Pr\{\Delta_i > c_i \quad \text{for} \quad 1 \le i \le j|\pi_0=\pi_1\}$$

$$\le \Pr\{\Delta_j < c_j|\pi_0=\pi_1\}$$

$$= \Pr\{|V_1(\underset{\sim}{X}_{0m_j};\underset{\sim}{X}_{1m_j})-V_2(\underset{\sim}{X}_{0m_j};\underset{\sim}{X}_{2m_j})| < c_j|\pi_0=\pi_1\}$$

$$= \Pr\{V_1-V_2<c_j,\ V_1>V_2|\pi_0=\pi_1\}+\Pr\{V_2-V_1<c_j,\ V_2>V_1|\pi_0=\pi_1\}$$

$$\le \Pr\{V_1-V_2 > 0|\pi_0=\pi_1\}+\Pr\{V_1-V_2 > -c_j|\pi_0=\pi_1\}$$

$$= \Pr\{V_1-V_2+\theta > \theta|\pi_0=\pi_1\}+\Pr\{V_1-V_2+\theta > \theta-c_j|\pi_0=\pi_1\}$$

$$\le 2\Pr\{V_1-V_2+\theta > c_j\}$$

for j sufficiently large. By following the exactly same argument as in Proposition 4.1, the right side of (4.20) tends to zero as $j \to \infty$. The proof is complete.

<u>Proposition 4.2'</u> Each of the PMC's of the rule (N', φ) is less than P.

<u>Proof.</u> Since (N', φ) terminates we can write

$$PMC_2 = \mathcal{E}(\varphi|\pi_0=\pi_2)$$

$$= \sum_{i=1}^{\infty} \Pr\{\Delta_j < c_j \quad \text{for} \quad j < i,\ \Delta_i \ge c_i,\ V_1 \le V_2|\pi_0=\pi_1\}$$

$$\le \sum \Pr\{V_2-V_1 \ge c_i|\pi_0=\pi_2\}$$

$$= \sum \Pr\{V_2-V_1+\theta \ge \theta+c_i|\pi_0=\pi_2\}$$

$$= \sum \Pr\{V_2-V_1+\theta \ge c_i|\pi_0=\pi_2\}$$

$$\le \sum \exp(-[m_i/m]c_i^2/2a^2) \le \sum \alpha_i \le p.$$

Similarly,

$$PMC_1 = \mathcal{E}(1-\varphi|\Pi_0=\Pi_1) \leq p.$$

In exactly the same way, we can prove that the moment generating function of N' exists, if the sequence $\{m_i\}$ is chosen suitably.

4.2. Sequential rules for classification into one of two multivariate normal populations.

For convenience, we shall follow Srivastava's notations. The problem is to classify $N_p(\mu_0, \Sigma)$ into one of $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$. When all the parameters are known and a sample of size n is taken from Π_0 the minimax rule is to classify Π_0 into Π_1 or Π_2 according as

(4.21) $\bar{X}_0' \Sigma^{-1}(\mu_1-\mu_2)-\frac{1}{2}(\mu_1+\mu_2)' \Sigma^{-1}(\mu_1-\mu_2) \gtrless 0,$

where $\bar{X}_0$ is the sample mean. The two PMC's given by e_{12} and e_{21}, are equal and their common value is given by

(4.22) $e_{12} = e_{21} = 1-\Phi(\frac{1}{2}n^{\frac{1}{2}}D),$

where $\Phi(x) = \int_{-\infty}^{x}(2\Pi)^{-\frac{1}{2}}\exp(-\frac{1}{2}y^2)dy,$

$$D^2 = \delta' \Sigma^{-1}\delta, \quad \text{and} \quad \delta = \mu_1-\mu_2.$$

To control probabilities of misclassification, Srivastava (1973) proposed sequential rules in the following two cases:

Case I. δ is known but Σ is unknown.

In this case we need to sample only from one of two populations, π_1 and π_2, which, without loss of generality, may be taken to be π_1. Given α, let $\Phi(a) = 1-\alpha$. Define

$$(4.23)\quad n\bar{X}_{in} = \sum_{j=1}^{n} X_{ij}, \qquad i=0,1.$$

and

$$(4.24)\quad mS_m = \sum_{i=0}^{1} \sum_{j=1}^{n} (X_{ij}-\bar{X}_{in})(X_{ij}-\bar{X}_{in})', \quad m = 2(n-1),$$

where $\{X_{ij}\}$ is a sequence of mutually independent random p-vector from $N_p(\mu_i, \Sigma)$, $i=0,1$. Then a stopping variable N is defined by

$$(4.25)\quad N = \text{the smallest integer } n(\geq n_0) \text{ such that}$$

$$n \geq 8a^2/\delta' S_m^{-1}\delta,$$

where $2n_0 \geq p+2$. When sampling is stopped at $N=n$, classify π_0 into π_1 or π_2 according as

$$(4.26)\quad (\bar{X}_{0n}-\bar{X}_{1n}+\tfrac{1}{2}\delta)' S_m^{-1}\delta \gtrless 0.$$

Define

$$(4.27)\quad Y_{ij} = \Sigma^{-\frac{1}{2}} X_{ij}, \quad n\bar{Y}_{in} = \sum_{j=1}^{n} Y_{ij}$$

$$mS_m^* = \sum_{i=0}^{1} \sum_{j=1}^{n} (Y_{ij}-\bar{Y}_{in})(Y_{ij}-\bar{Y}_{in})' = \Sigma^{-\frac{1}{2}}(mS_m)\,\Sigma^{-\frac{1}{2}}.$$

Then

$$(4.28)\quad \delta' S_m^{-1}\delta = \delta'(\Sigma^{\frac{1}{2}} S_m^* \Sigma^{\frac{1}{2}})^{-1}\delta$$

$$= (\Sigma^{-\frac{1}{2}}\delta)' S_m^{*-1}(\Sigma^{-\frac{1}{2}}\delta) = \delta^{*\prime} S_m^{*-1}\delta^*,$$

where $\delta^* = \Sigma^{-\frac{1}{2}}\delta$. Note that mS_m^* is distributed as $W_p(m,I)$.

Now we shall obtain some asymptotic properties of this sequential rule as $D \to 0$ (or $\delta^* \to 0$). From (4.25) and (4.28),

$$N = N(\delta^*) = \text{the smallest } n \geq n_0 \text{ such that}$$

$$n \geq 8a^2/\delta^{*\prime} S_m^{-1}\delta^*$$

or

$$(4.29)\quad (1/m)\delta^{*\prime}\delta^*/\delta^{*\prime}(mS_m^*)^{-1}\delta^* \leq n\delta^{*\prime}\delta^*/8a^2$$

Since for $\delta^* \neq 0$, $\delta^{*\prime}\delta^*/\delta^{*\prime}(mS_m^*)^{-1}\delta^* \sim \chi^2_{m-p+1}$

$$(4.30)\quad (1/m)\delta^{*\prime}\delta^*/\delta^{*\prime}(mS_m^*)^{-1}\delta^* \to 1 \text{ a.s. as } n \to \infty$$

From (4.29) and Lemma 1. of Chow and Robbins (1965), we have the following.

<u>Lemma 4.1.</u> (i) $N \to \infty$ a.s. as $\delta^* \to 0$.

(ii) $N/(8a^2/\delta^{*\prime}\delta^*) \to 1$ a.s. as $\delta^* \to 0$.

Note that the rule is now studied in terms of Y_{ij}'s and the a.s. convergence as $\delta^* \to 0$ is meaningful (contrary to Srivastava's development).

Lemma 4.2. (Asymptotic efficiency) $\lim_{\delta^*\to 0} \mathcal{E}N/(8a^2/\delta^{*\prime}\delta^*) = 1$

Proof. It is enough to show that $\{N\delta^{*\prime}\delta^*\}_{\delta^{*\prime}\delta^* > 0}$ is uniformly integrable. According to a result of Bickel and Yahav (Lemma 3.2, 1968), it is sufficient to prove that $\sum_{k=1}^{\infty} \sup_{0<\delta^{*\prime}\delta^*<\epsilon} \Pr\{N\delta^{*\prime}\delta^* > k\} < \infty$ for some $\epsilon > 0$. Now, for $0 < \delta^{*\prime}\delta^* < \epsilon$

$$\begin{aligned}
\Pr\{N\delta^{*\prime}\delta^* > k\} &= \Pr\{N > k/\delta^{*\prime}\delta^*\} \\
&\le \Pr\{N > k(\delta^*)\}, \quad \text{where} \quad k(\delta^*) = [k/\delta^{*\prime}\delta^*] \\
&\le \Pr\{k(\delta^*) < 8a^2/\delta^{*\prime}S_f^{*-1}\delta^*\}, \quad \text{where} \quad f = 2(k(\delta^*)-1) \\
&= \Pr\{\delta^{*\prime}\delta^*/\delta^{*\prime}(fS_f^*)^{-1}\delta^* > f\delta^{*\prime}\delta^* k(\delta^*)/8a^2\} \\
&\le (64a^4/k^2(\delta^*)(\delta^{*\prime}\delta^*)^2)\cdot\mathcal{E}(\chi^2_{f-p+1})^2/f^2 \\
&\le (64a^4/(k-\delta^{*\prime}\delta^*)^2)\cdot(f-p+1)(f-p+3)/f^2 \\
&\le 64a^4/(k-\epsilon)^2
\end{aligned}$$

for ϵ sufficiently small.

Hence

$$\sum_{k=1}^{\infty} \sup_{0<\delta^{*\prime}\delta^*<\epsilon} \Pr\{N\delta^{*\prime}\delta^* > k\} < \infty,$$

which completes the proof.

Let $e_{ij} = \Pr\{\text{classifying } \Pi_0 \text{ into } \Pi_j | \Pi_0 = \Pi_i\}$. Then

Theorem 4.1. $\lim_{\delta^*\to 0} e_{12} = \lim_{\delta^*\to 0} e_{21} = \alpha.$

<u>Proof.</u> We shall only prove for e_{12}. Let $M = 2(N-1)$, we have

$$(4.31)\quad e_{12} = \Pr\{(N/2)^{\frac{1}{2}}(\bar{X}_{ON}-\bar{X}_{1N})'S_M^{-1}\delta < -\tfrac{1}{2}(N/2)^{\frac{1}{2}}\delta' S_M^{-1}\delta \,|\, \Pi_0=\Pi_1\}$$

$$= \Pr\{(N/2)^{\frac{1}{2}}(\bar{X}_{ON}-\bar{X}_{1N})'S_M^{-1}\delta/(\delta' S_M^{-1}\Sigma\, S_M^{-1}\delta)^{\frac{1}{2}}$$

$$< -\tfrac{1}{2}(N/2)^{\frac{1}{2}}\delta' S_M^{-1}\delta/(\delta' S_M^{-1}\Sigma\, S_M^{-1}\delta)^{\frac{1}{2}} \,|\, \Pi_0=\Pi_1\}$$

$$= 1-\mathcal{E}\Phi[(N/8)^{\frac{1}{2}}\delta^{*\prime} S_M^{*-1}\delta^*/(\delta^{*\prime} S_M^{*-2}\delta^*)^{\frac{1}{2}}]$$

$$= 1-\mathcal{E}\Phi[a\cdot(N/8a^2/\delta^{*\prime}\delta^*))^{\frac{1}{2}}\cdot(\delta^{*\prime} S_M^{*-1}\delta^*/\delta^{*\prime}\delta^*)\cdot$$

$$(\delta^{*\prime}\delta^*/\delta^* S_M^{*-2}\delta^*)^{\frac{1}{2}}]$$

Now for any orthogonal matrix L we can write

$$(4.32)\quad \delta^{*\prime} S_M^{*-2}\delta^* = \delta^* L' L S_M^{*-1} L' L S_M^{*-1} L' L\delta^*$$

$$= (L\delta^*)' A^{-1} A^{-1}(L\delta^*), \qquad A = L S_m^{*-1} L'.$$

If we choose L with first row as $\delta^{*\prime}/(\delta^{*\prime}\delta^*)^{\frac{1}{2}}$, then

$$(4.33)\quad \delta^{*\prime} S_M^{*-2}\delta^* = \delta^{*\prime}\delta^*(\text{first row of } A^{-1})(\text{first column of } A^{-1})$$

Therefore,

$$\delta^{*\prime} S_M^{*-2}\delta^*/\delta^{*\prime}\delta^* = (\text{first column of } A^{-1})(\text{first column of } A^{-1})$$

Since $MA \sim W_p(M,I)$, $A \to I_p$ a.s. as $\delta^* \to 0$

Hence

$$\delta^{*\prime} S_M^{*-2}\delta^*/\delta^{*\prime}\delta^* \to 1 \text{ a.s. as } \delta^* \to 0.$$

Also we have seen from (4.30) and Lemma 4.1 that

$N/(8a^2/\delta^{*\prime}\delta^*) \to 1$ and $\delta^{*\prime}S_M^{*-1}\delta^*/\delta^{*\prime}\delta^* \to 1$ a.s. as $\delta^* \to 0$.

By the dominated convergence theorem, from (4.31) we have

$$\lim_{\delta^*\to 0} e_{12} = 1-\Phi(a) = \alpha.$$

Case II. Both δ and Σ are unknown.

Now sampling is carried out sequentially from Π_2 as well.

Let

$$(4.34)\quad tW_t = \sum_{i=0}^{2}\sum_{j=1}^{n}(X_{ij}-\bar{X}_{in})(X_{ij}-\bar{X}_{in})', \quad t = 3(n-1),$$

where $n\bar{X}_{2n} = \sum_{j=1}^{n} X_{2j}$. Then the sampling rule is

(4.35) N' = the smallest integer $n(\geq n_0)$ such that

$$n \geq 6a^2/\delta_n' W_t^{-1}\delta_n;$$

where $\delta_n = \bar{X}_{1n}-\bar{X}_{2n}$. When the sampling is stopped at $N' = n$, classify Π_0 into Π_1 or Π_2 according as

$$(4.36)\quad [\bar{X}_{0n}-\tfrac{1}{2}(\bar{X}_{1n}+\bar{X}_{2n})]'W_t^{-1}\delta_n \gtrless 0.$$

Let

$$(4.37)\quad Z_{ij} = \Sigma^{-\frac{1}{2}}(X_{ij}-\mu_i), \quad nZ_{in} = \sum_{j=1}^{n} Z_{ij},$$

$$tW_t^* = \sum_{i=0}^{2}\sum_{j=1}^{n}(Z_{ij}-\overline{Z}_{in})(Z_{ij}-\overline{Z}_{in})' = \Sigma^{-\frac{1}{2}}(tW_t)\,\Sigma^{-\frac{1}{2}},$$

$$\delta_n^* = \overline{Z}_{1n}-\overline{Z}_{2n} = \Sigma^{-\frac{1}{2}}(\overline{X}_{1n}-\overline{X}_{2n})-\Sigma^{-\frac{1}{2}}(\mu_1-\mu_2) = \Sigma^{-\frac{1}{2}}\delta_n-\delta^*$$

Then

$$(4.38)\quad \delta_n' W_t^{-1}\delta_n = (\delta_n^*+\delta^*)' W_t^{*-1}(\delta_n^*+\delta^*)$$

$$= (n/2)(\delta_n^*+\delta^*)'(tW_t^*)^{-1}(\delta_n^*+\delta^*)(t/(n/2))$$

$$= (U/V)(t/(n/2)),$$

where U and V are mutually independent and $U \sim \chi_p^2((n/2)\delta^{*\prime}\delta^*)$, $V \sim \chi^2_{t-p+1}$ (see Anderson Theorem 5.2.2, p. 106, 1958).

Note that U/V is stochastically larger than χ_p^2/V. Therefore, if we define

N^* = the first $n(\geq n_0)$ such that

$$n \geq 6a^2/[(\chi_p^2/V)(t/(n/2))],$$

then N^* is stochastically larger than N', and N^* is independent of δ^*. It is clear that there is positive probability that N^* is finite. Hence it is not true that

$$\lim_{\delta^*\to 0} N' = \infty \quad \text{a.s.},$$

which is the error in Srivastava's argument.

APPENDIX

DERIVATIONS TO OBTAIN A TAYLOR SERIES EXPANSION OF THE FUNCTION $\partial\Psi(\quad)$

Let

(A.1) $\varphi_2(x,y;r) = (1/2\pi)\ (1-r^2)^{-\frac{1}{2}} e^{-\frac{1}{2}(x^2-2rxy+y^2)/(1-r^2)} \quad |r| < |$

(A.2) $\Psi(u,v;r) = \int_v^\infty \int_u^\infty \varphi_2(x,y;r)dxdy.$

Then

(A.3) $\partial\Psi/\partial u = -\int_v^\infty (1/2\pi)\ (1-r^2)^{-\frac{1}{2}} e^{-\frac{1}{2}(u^2-2ruy+y^2)/(1-r^2)} dy$.

$= -\int_v^\infty (1/2\pi)\ (1-r^2)^{-\frac{1}{2}} e^{-\frac{1}{2}(y^2-2ruy+u^2r^2+u^2(1-r^2))/(1-r^2)} dy$

$= -\int_v^\infty (2\pi(1-r^2))^{-\frac{1}{2}} e^{-\frac{1}{2}(y-ur)^2/(1-r^2)}\ (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2} dy$

$= -\varphi_1(u)[1-\Phi_1((v-ur)/(1-r^2)^{\frac{1}{2}})]$

where

(A.4) $\varphi_1(u) = (2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}u^2}$

(A.5) $\Phi_1(u) = \int_{-\infty}^u \varphi_1(x)dx$

Similarly,

(A.6) $\partial\Psi/\partial v = -\varphi_1(v)[1-\Phi_1((u-vr)/(1-r^2)^{\frac{1}{2}})]$

and for $o < |r| < 1$

(A.7) $$\partial\Psi/\partial r = \int_v^\infty\int_u^\infty \partial/\partial r(\varphi_2(x,y;r))dxdy$$

$$= \int_v^\infty\int_u^\infty \partial^2/\partial x\partial y(\varphi_2(x,y;r))dxdy$$

$$= \varphi_2(u,v;r)$$

(A.8) $$\partial^2\Psi/\partial u^2 = \partial/\partial u(-\varphi_1(u)[1-\Phi_1((v-ur)/(1-r^2)^{\frac{1}{2}})])$$

$$= (1-\Phi_1((v-ur)/(1-r^2)^{\frac{1}{2}}))\varphi_1(u)u+\varphi_1(u)\varphi_1((v-ur)/$$

$$(1-r^2))\ (-r/(1-r^2)^{\frac{1}{2}})$$

$$= \varphi_1(u)[-\varphi_1((v-ur)/(1-r^2)^{\frac{1}{2}}+u(1-\Phi_1((v-ur)/(1-r^2)^{\frac{1}{2}}))]$$

(A.9) $$\partial^2\Psi/\partial v^2 = \varphi_1(v)[-\varphi_1((u-vr)/(1-r^2)^{\frac{1}{2}})r/(1-r^2)^{\frac{1}{2}}+v(1-\Phi_1((u-vr)/$$

$$(1-r^2)))]$$

(A.10) $$\partial^2\Psi/\partial r^2 = \partial/\partial r(\varphi_2(u,v;r)) = \partial^2/\partial v\partial u(\varphi_2(v,u;r))$$

$$= \partial/\partial v(-\varphi_2(u,v;r)(u-vr)/(1-r^2))$$

$$= \varphi_2(v,v;r)[r+(u-vr)(v-ur)/(1-r^2)]/(1-r^2)$$

$$(A.11)\ \partial^2\Psi/\partial v\partial u = \varphi_1(u)\varphi_1((v-ur)/(1-r^2)^{\frac{1}{2}})/(1-r^2)^{\frac{1}{2}}$$

$$(A.12)\ \partial^2\Psi/\partial r\partial u = \varphi_1(u)\varphi_1((v-ur)/(1-r^2)^{\frac{1}{2}})(-u+vr)/(1-r^2)^{3/2}$$

$$(A.13)\ \partial^2\Psi/\partial u\partial v = \varphi_1(v)\varphi_1((u-vr)/(1-r^2)^{\frac{1}{2}})/(1-r^2)^{\frac{1}{2}}$$

$$(A.14)\ \partial^2\Psi/\partial r\partial v = \varphi_1(v)\varphi_1((u-vr)/(1-r^2)^{\frac{1}{2}})(-u+vr)/(1-r^2)^{3/2}$$

$$(A.15)\ \partial^2\Psi/\partial u\partial r = -\varphi_2(u,v;r)(u-vr)/(1-r^2)$$

$$(A.16)\ \partial^2\Psi/\partial v\partial r = -\varphi_2(u,v;r)(v-ur)/(1-r^2)$$

Then (as Taylor Series expansion) for real a, b, and $|\rho| < 1$ we have

$$(A.17)\ \Psi(u,v;r) = \Psi(a,b;\rho)+[-\varphi_1(a)(1-\Phi_1[(b-a\rho)/(1-\rho^2)^{\frac{1}{2}}))](u-a)$$

$$+[-\varphi_1(b)(1-\Phi_1[(a-b\rho)/(1-\rho^2)^{\frac{1}{2}}))](v-b)+\varphi_2(a,b;\rho)\times$$

$$(r-\rho)+\tfrac{1}{2}\varphi_1(a)[-\varphi_1((b-a\rho)/(1-\rho^2)^{\frac{1}{2}})\rho/(1-\rho^2)^{\frac{1}{2}}$$

$$+a(1-\Phi_1((b-a\rho)/(1-\rho^2)^{\frac{1}{2}}))](u-a)^2+\tfrac{1}{2}\varphi_1(b)\times$$

$$[-\varphi_1((a-b\rho)/(1-\rho^2)^{\frac{1}{2}})\rho/(1-\rho^2)^{\frac{1}{2}}+b(1-\Phi_1((a-b\rho)/$$

$$(1-\rho^2)^{\frac{1}{2}}))](v-b)^2+\tfrac{1}{2}\varphi_2(a,b;\rho)[\rho+(a-b\rho)(b-a\rho)/$$

$$(1-\rho^2)](r-\rho)^2/(1-\rho^2)+\tfrac{1}{2}[\varphi_1(a)\varphi_1((b-a\rho)/(1-\rho^2)^{\frac{1}{2}})$$

$$+\varphi_1(b)\varphi_1((a-b\rho)/(1-\rho^2)^{\frac{1}{2}})]\ (u-a)(v-b)/(1-\rho^2)^{\frac{1}{2}}$$

$$+\tfrac{1}{2}[\varphi_1(a)\varphi_1((b-a\rho)/(1-\rho^2)^{\frac{1}{2}})(-a+b\rho)/(1-\rho^2)^{3/2}-$$

$$\varphi_2(a,b;\rho)\ (a-b\rho)/(1-\rho^2)](u-a)(r-\rho)+\tfrac{1}{2}[\varphi_1(b)$$

$$\varphi_1((a-b\rho)/(1-\rho^2)^{\frac{1}{2}})(-b+a\rho)/(1-\rho^2)^{3/2}-\varphi_2(a,b;\rho)\times$$

$$(b-a\rho)/(1-\rho^2)](v-b)(r-\rho)+R$$

where R is a remainder term. Simplifying (A.17) we have

$$\text{(A.18)}\ \Psi(u,v;r) = \Psi(a,b;\rho)+\varphi_1(a)[1-\Phi_1((b-a\rho)/(1-\rho^2)^{\frac{1}{2}}]\ [\tfrac{1}{2}a(u-a)^2-$$

$$(u-a)]+\varphi_1(b)[1-\Phi_1((a-b\rho)/(1-\rho^2)^{\frac{1}{2}}][\tfrac{1}{2}b(v-b)^2-$$

$$(v-b)]+\varphi_1(a)\varphi_1((b-a\rho)/(1-\rho^2)^{\frac{1}{2}})[-\tfrac{1}{2}\rho(u-a)^2+\tfrac{1}{2}(u-a)\times$$

$$(v-b)+\tfrac{1}{2}(a-b\rho)(u-a)(r-\rho)/(1-\rho^2)]/(1-\rho^2)^{\frac{1}{2}}+\varphi_1(b)\times$$

$$\varphi_1((a-b\rho)/(1-\rho^2)^{\frac{1}{2}})\ [-\tfrac{1}{2}\rho(v-b)^2+\tfrac{1}{2}(u-a)(v-b)-$$

$$\tfrac{1}{2}(b-a\rho)(v-b)(r-\rho)/(1-\rho^2)]/(1-\rho^2)^{\frac{1}{2}}+\varphi_2(a,b;\rho)\times$$

$$[(r-\rho)+\tfrac{1}{2}(r-\rho)^2(\rho+(a-b\rho)(b-a\rho)/(1-\rho^2))/(1-\rho^2)$$

$$-\tfrac{1}{2}(a-b\rho(u-a)(r-\rho)/(1-\rho^2)-\tfrac{1}{2}(b-a\rho)(v-b)(r-\rho)/$$

$$(1-\rho^2)]+R$$

For $a = \frac{1}{2}\alpha^{\frac{1}{2}}_{22}$, $b = \frac{1}{2}(\alpha_{22}-\alpha_{33})/(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{\frac{1}{2}}$,

$\rho = (\alpha_{22}-\alpha_{23})/([\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})]^{\frac{1}{2}}$ and $|\rho| < 1$

we have

(A.19) $(b-a\rho)/(1-\rho^2)^{\frac{1}{2}} = \frac{1}{2}\alpha^{\frac{1}{2}}_{22}(\alpha_{23}-\alpha_{33})/(\alpha_{22}\alpha_{33}-\alpha^2_{23})^{\frac{1}{2}}$

(A.20) $(a-b\rho)/(1-\rho^2)^{\frac{1}{2}} = \frac{1}{2}(2\alpha_{22}\alpha_{33}-\alpha_{22}\alpha_{23}-\alpha_{33}\alpha_{23})/$

$[(\alpha_{22}+\alpha_{33}-2\alpha_{23})(\alpha_{22}\alpha_{33}-\alpha^2_{23})]^{\frac{1}{2}}$

(A.21) $(1-\rho^2)^{-\frac{1}{2}} = \alpha^{\frac{1}{2}}_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{\frac{1}{2}}/(\alpha_{22}\alpha_{33}-\alpha^2_{23})^{\frac{1}{2}}$.

(A.22) $\rho/(1-\rho^2)^{\frac{1}{2}} = (\alpha_{22}-\alpha_{23})/(\alpha_{22}\alpha_{33}-\alpha^2_{23})^{\frac{1}{2}}$

(A.23) $(b-a\rho)/(1-\rho^2) = \frac{1}{2}\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{\frac{1}{2}}(\alpha_{23}-\alpha_{33})/$

$(\alpha_{22}\alpha_{33}-\alpha^2_{23})$

(A.24) $(a-b\rho)/(1-\rho^2) = \frac{1}{2}\alpha^{\frac{1}{2}}_{22}(2\alpha_{22}\alpha_{33}-\alpha_{22}\alpha_{23}-\alpha_{33}\alpha_{23})/$

$(\alpha_{22}\alpha_{33}-\alpha^2_{23})$

(A.25) $(a-b\rho)(b-a\rho)/(1-\rho^2) = (1/4)\alpha^{\frac{1}{2}}_{22}(\alpha_{23}-\alpha_{133})(2\alpha_{22}\alpha_{33}$

$-\alpha_{22}\alpha_{23}-\alpha_{33}\alpha_{23})/(\alpha_{22}+\alpha_{33}$

$-2\alpha_{23})^{\frac{1}{2}}(\alpha_{22}\alpha_{33}-\alpha^2_{23})$

$$(A.26)\quad (\rho+(a-b\rho)(b-a\rho)/(1-\rho^2)) = [\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{\frac{1}{2}}/$$

$$(\alpha_{22}\alpha_{33}-\alpha^2_{23})]\ [(\alpha_{22}-\alpha_{23})/\alpha^{\frac{1}{2}}_{22}+(1/4)\alpha^{\frac{1}{2}}_{22}$$

$$(\alpha_{23}-\alpha_{33})(2\alpha_{22}\alpha_{33}-\alpha_{22}\alpha_{23}-\alpha_{33}\alpha_{23})/$$

$$(\alpha_{22}\alpha_{33}-\alpha^2_{23})]$$

Next we are going to compute some expectations.

Consider the matrix V defined by

$$(A.27)\quad (1/m)A = I+(1/m^{\frac{1}{2}})V, \quad \text{where } A \sim W_p(I,m)$$

and for fixed η_2, η_3 we define

$$(A.28)\quad \alpha_{22} = \eta_2'\eta_2, \quad \alpha_{33} = \eta_3'\eta_3, \quad \alpha_{23} = \eta_2'\eta_3$$

Lemma A.1 $\quad \mathcal{E}V^2 = (p+1)I$

Proof. By definition of V,

$$V = m^{\frac{1}{2}}((1/m)A-I)$$

$$= (1/m^{\frac{1}{2}})(A-mI)$$

$$V^2 = (1/m)(A^2-2mA+m^2I)$$

We can write A as $A = \sum_{i=1}^{m} Z_i Z_i'$, where Z_i's are identically independently distributed as $N_p(0,I)$.

Then

$$A^2 = (\sum_{i=1}^{m} z_i z_i')(\sum_{j=1}^{m} z_j z_j') = \sum_{i=1}^{m} z_i z_i' z_i z_i' + \sum_{i \neq j} z_i z_i' z_j z_j'$$

$$\mathcal{E}A^2 = m(p+2)I + m(m-1)I = m(m+p+1)I$$

$$\mathcal{E}V^2 = (1/m)(m(m+p+1)I - 2m^2 I + m^2 I)$$

$$= (p+1)I$$

<u>Lemma A.2</u> (i) $\mathcal{E}(\eta_2' V \eta_2)^2 = 2(\eta_2'\eta_2)^2 = 2\alpha^2{}_{22}$

(ii) $\mathcal{E}(\eta_3' V \eta_3)^2 = 2\alpha^2{}_{33}$

<u>Proof.</u> By definition,

$$V = (1/m^{\frac{1}{2}})(A - mI)$$

$$(\eta_2' V \eta_2)^2 = (1/m)(\eta_2' A \eta_2 - m\eta_2'\eta_2)^2$$

$$= (1/m)(\eta_2'\eta_2)^2(\eta_2' A \eta_2/\eta_2'\eta_2 - m)^2$$

Since $\eta_2' A \eta_2/\eta_2'\eta_2 \sim \chi^2 m$ we have

$$\mathcal{E}(\eta_2' V \eta_2)^2 = (1/m)(\eta_2'\eta_2)^2 \cdot 2m = 2(\eta_2'\eta_2)^2 = 2\alpha^2{}_{22}$$

<u>Lemma A.3</u> $\mathcal{E}\eta_2' V \eta_2 \eta_3' V \eta_3 = 2(\eta_2'\eta_3)^2 = 2\alpha^2{}_{23}$

<u>Proof.</u> We can express η_3 as

$$\eta_3 = k\eta_2 + e$$

where $k\eta_2$ is the orthogonal projection of η_3 on $\varphi(\eta_2)$, the space generated by η_2, and e is the perpendicular of η_3 with respect to $\varphi(\eta_2)$. The coefficient k is known as $\eta_2'\eta_3/\eta_2'\eta_2$. Then

$$\text{(A.29)}\quad \eta_2'V\eta_2\eta_3'V\eta_3 = \eta_2'V\eta_2(k\eta_2+e)'(k\eta_2+e)$$

$$= k^2(\eta_2'V\eta_2)^2+2k\eta_2'V\eta_2\eta_2'Ve+\eta_2'V\eta_e e'Ve$$

Now $V = (1/m^{\frac{1}{2}})(A-mI)$, $A \sim W_p(I,m)$.
Since η_2 and e are mutually orthogonal we can choose an orthogonal matrix L with first row as $\eta_2'/(\eta_2'\eta_2)^{\frac{1}{2}}$, second row as $e'/(e'e)^{\frac{1}{2}}$, and define A^* as

$A^* = LAL'$, A^* is again distributed as $W_p(I,m)$. Then the (1,1)th, (1,2)th, (2,2)th elements of A^* are

$$\text{(A.30)}\quad a_{11}^* = \eta_2'A\eta_2/(\eta_2'\eta_2),\quad a_{12}^* = \eta_2'Ae/[\eta_2'\eta_2 e'e]^{\frac{1}{2}},$$

$$a_{22}^* = e'Ae/(e'e)$$

a_{11}^* and a_{22}^* are independent because $A^* \sim W_p(I,m)$. Therefore $\eta_2'V\eta_2$ and $e'Ve$. Consequently,

$$\mathcal{E}\eta_2'V\eta_2 e'Ve = 0.$$

Furthermore, we can write

$$a_{11}^* = \sum_{i=1}^{m} z_{i1}^2,\ a_{12}^* = \sum_{i=1}^{m} z_{i1}z_{i2},\quad \text{where } z_{ij}\text{'s are}$$

identically independently distributed as $N(0,1)$, $i = 1,\ldots, m$, $j = 1,2,\ldots.$

$$a^*_{i1}a^*_{i2} = (\sum_{i=1}^{m} z_{i1}^2)(\sum_{i=1}^{m} z_{i1}z_{i2}) = \sum_{i=1}^{m} z_{i1}^3 z_{i2} + \sum_{i \neq j} z_{i1}^2 z_{j1} z_{j2}$$

$\mathcal{E}a^*_{i1}a^*_{i2} = 0$, which entails $\mathcal{E}\eta_2' A\eta_2\eta_2' Ae = 0$. Hence

$$\text{(A.31)} \quad \mathcal{E}\eta_2' v\eta_2\eta_2' ve = \mathcal{E}\eta_2'((1/m^{\frac{1}{2}})(A-mI))\eta_2\eta_2'((1/m^{\frac{1}{2}})(A-mI))e$$

$$= (1/m)\mathcal{E}(\eta_2' A\eta_2\eta_2' Ae - m\eta_2'\eta_2\eta_2' Ae)$$

$$= 0$$

because $\mathcal{E}A = I$, and $\eta_2' e = 0$. Thus from (A.29), we have

$$\mathcal{E}\eta_2' v\eta_2\eta_3' v\eta_3 = k^2\mathcal{E}(\eta_2' v\eta_2)^2 = k^2\, 2(\eta_2'\eta_2)^2 = (\eta_2'\eta_3/\eta_2'\eta_2)^2.$$

$$2(\eta_2'\eta_2)^2 = 2(\eta_2'\eta_3)^2 = 2\alpha^2_{23}$$

<u>Lemma A.4</u> (i) $\mathcal{E}\eta_2' v\eta_2\eta_2' v\eta_3 = 2(\eta_2'\eta_2)(\eta_2'\eta_3) = 2\alpha_{22}\alpha_{23}$

(ii) $\mathcal{E}\eta_2' v\eta_3\eta_3' v\eta_3 = 2(\eta_2'\eta_3)(\eta_3'\eta_3) = 2\alpha_{33}\alpha_{23}$

<u>Proof.</u> As in Lemma A.3 we can write

$$\eta_2' v\eta_2\eta_2' v\eta_3 = \eta_2' v\eta_2\eta_2' v(k\eta_2+e) = k(\eta_2' v\eta_2)^2+\eta_2' v\eta_2\eta_2' ve$$

$$\mathcal{E}\eta_2' v\eta_2\eta_2' v\eta_3 = k\cdot 2(\eta_2'\eta_2)^2+0 = (\eta_2'\eta_3/\eta_2'\eta_2)\, 2(\eta_2'\eta_2)^2$$

$$= 2(\eta_2'\eta_2)(\eta_2'\eta_3) = 2\alpha_{22}\alpha_{23}$$

<u>Lemma A.5</u> $\quad \mathcal{E}(\eta_2' v\eta_3)^2 = (\eta_2'\eta_3)^2+(\eta_2'\eta_2)(\eta_3'\eta_3) = \alpha^2_{23}+\alpha_{22}\alpha_{33}$

<u>Proof.</u> $\quad (\eta_2' v\eta_3)^2 = [\eta_2' v(k\eta_2+e)]^2 = (k\eta_2' v\eta_2+\eta_2' ve)^2$

$$= k^2(\eta_2' v\eta_2)^2+2k\eta_2' v\eta_2\eta_2' ve+(\eta_2' ve)^2$$

$$\eta_2' ve = m^{-\frac{1}{2}}\eta_2'(A-mI)e = m^{-\frac{1}{2}}\eta_2' Ae$$

$$= m^{-\frac{1}{2}}\, a^*_{12}(\eta_2'\eta_2)^{\frac{1}{2}}(e'e)^{\frac{1}{2}} \qquad \text{(see A.30)}$$

$$= m^{-\frac{1}{2}}(\eta_2'\eta_2)^{\frac{1}{2}}(e'e)^{\frac{1}{2}}\left(\sum_{i=1}^{m} z_{i1}z_{i2}\right),\ z_{ij} \sim N(0.1)$$

and

$$e'e = (\eta_3-k\eta_2)'(\eta_3-k\eta_2) = \eta_3'\eta_3-(\eta_2'-\eta_3)^2/\eta_2'\eta_2$$

Hence

$$\mathcal{E}(\eta_2' ve)^2 = (1/m)\eta_2'\eta_2(\eta_3'\eta_3-\eta_2'\eta_3/\eta_2'\eta_2)\mathcal{E}\left(\sum_{i=1}^{m} z^2_{i1}z^2_{i2}+\right.$$

$$\left.\sum_{i\neq j} z_{i1}z_{i2}z_{j1}z_{j2}\right)$$

$$= \eta_2'\eta_2(\eta_3'\eta_3-(\eta_2'\eta_3)^2/\eta_2'\eta_2) = \eta_2'\eta_2\eta_3'\eta_3-(\eta_2'\eta_3)^2$$

Thus

$$\mathcal{E}(\eta_2' V\eta_3)^2 = k^2\mathcal{E}(\eta_2' V\eta_2)^2+\mathcal{E}(\eta_2' Ve)^2$$

$$= (\eta_2'\eta_3/\eta_2'\eta_2)^2\ 2(\eta_2'\eta_2)^2+\eta_2'\eta_2\eta_3'\eta_3-(\eta_2'\eta_3)^2$$

$$= \eta_2'\eta_2\eta_3'\eta_3+(\eta_2'\eta_3)^2 = \alpha^2_{23}+\alpha_{.22}\alpha_{.33}$$

From the above lemmas and the fact that Y_1, Y_2, Y_3, and V are independent and $Y_i \sim N_p(0,(m/n)I)$ $i = 1, 2, 3\ldots$, $\mathcal{E}V = 0$. Then for sufficiently large m, n and $m/n \doteq 3$, we are now able to compute the expectations of the functions defined in Chapter One.

(A.32) $\mathcal{E}D(Y_1,Y_2,V) = -\frac{1}{2}(p-1)(\frac{1}{2}\alpha^{\frac{1}{2}}_{22}+3\alpha^{-\frac{1}{2}}_{22})$

(A.33) $\mathcal{E}G(Y_2,Y_3,V) = -\frac{1}{2}(p-1)(\alpha_{22}-\alpha_{33})(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{-\frac{1}{2}} \times$

$$(\tfrac{1}{2}+3/(\alpha_{22}+\alpha_{33}-2\alpha_{23}))$$

(A.34) $\mathcal{E}K(Y_1,Y_2,Y_3,V) = (\alpha_{22}(\alpha_{.22}+\alpha_{33}-2\alpha_{23}))^{-\frac{1}{2}}[3(p-2)-3(p-1)\times$

$$(\alpha_{.22}-\alpha_{23})/\alpha_{22}-3(p-1)(\alpha_{22}-\alpha_{23})/$$

$$(\alpha_{22}+\alpha_{33}-2\alpha_{23})-2(\alpha_{.22}-\alpha_{.23})+$$

$$(\alpha_{.22}-\alpha_{.23})^2(3+2(\alpha_{.22}-\alpha_{.23}))/$$

$$(\alpha_{.22}(\alpha_{.22}+\alpha_{33}-2\alpha_{23}))]$$

(A.35) $\mathcal{E}C^2(Y_1,Y_2) = 3/2$

(A.36) $\mathcal{E}F^2(Y_1,Y_2,V) = [(\alpha_{22}\alpha_{33}-\alpha_{23})^2+3(\alpha_{22}+\alpha_{33})]/$

$(\alpha_{22}+\alpha_{33}-2\alpha_{23})-(3/2)(\alpha_{22}-\alpha_{33})^2/$

$(\alpha_{22}+\alpha_{33}-2\alpha_{23})^2$

(A.37) $\mathcal{E}H^2(Y_1,Y_2,Y_3,V) = 2[-4(\alpha_{22}-\alpha^2_{23})-3(\alpha_{22}-\alpha_{23})$

$-3(\alpha_{22}-\alpha_{23})^2/\alpha_{22}-3(\alpha_{22}-\alpha_{23})^2/$

$(\alpha_{22}+\alpha_{33}-2\alpha_{23})+3(\alpha_{22}+\alpha_{33}-2\alpha_{23})$

$+3\alpha_{22}+2\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})$

$-2(\alpha_{22}-\alpha_{23})^3(3+2(\alpha_{22}-\alpha_{23}))/$

$(\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23}))]/(\alpha_{22}$

$(\alpha_{22}+\alpha_{33}-2\alpha_{23}))$

(A.38) $\mathcal{E}C(Y_1,Y_2)F(Y_2,Y_3,V) = (3/4)[2\alpha_{22}-(\alpha_{22}-\alpha_{33})(\alpha_{22}-\alpha_{23})/$

$(\alpha_{22}+\alpha_{33}-2\alpha_{23})]/[\alpha_{22}$

$(\alpha_{22}+\alpha_{33}-2\alpha_{23})]^{\frac{1}{2}}$

(A.39) $\mathcal{E}C(Y_1,Y_2)H(Y_1,Y_2,Y_3,V) = (3/2)(\alpha_{.22}\alpha_{.33}-\alpha^2_{.23})/$

$[\alpha_{.22}(\alpha_{.22}+\alpha_{.33}-2\alpha_{.23})]^{3/2}$

(A.40) $\mathcal{E}F(Y_2,Y_3,V)H(Y_1,Y_2,Y_3,V)$

$= \frac{1}{2}(\alpha_{.22}\alpha_{.33}-\alpha^2_{23})(11\alpha_{.22}+5\alpha_{.33}-4\alpha_{.23})/[\alpha^{3/2}_{.22}(\alpha_{.22}\alpha_{.33}$

$-2\alpha_{23})^2]$

(A.41) $\mathcal{E}D^*(Y_1,Y_2) = (1/8)(2p+1)\alpha^{\frac{1}{2}}_{22}+3(p-1)/2\alpha^{\frac{1}{2}}_{.22}$

(A.42) $\mathcal{E}G^*(Y_1,Y_2,Y_3,V) = (\alpha_{.22}-\alpha_{.33})[-(2p+1)/8-(3/2)(p-1)/$

$(\alpha_{22}+\alpha_{33}-2\alpha_{.23})]/(\alpha_{.22}+\alpha_{.33}-2\alpha_{.23})^{\frac{1}{2}}$

(A.43) $\mathcal{E}K^*(Y_1,Y_2,Y_3,V) = [\alpha_{.22}(\alpha_{.22}+\alpha_{33}-2\alpha_{23})]^{-\frac{1}{2}}[-12-3(p-1)\times$

$(\alpha_{.22}-\alpha_{23})/(\alpha_{22}+\alpha_{.33}-2\alpha_{.23})-$

$(3(p-1)(\alpha_{.22}-\alpha_{.23})+6\alpha_{23})/\alpha_{.22}+$

$(\alpha_{.22}-\alpha_{23})^2(\alpha_{22}-\alpha_{23}+6)/[\alpha_{22}$

$(\alpha_{.22}+\alpha_{33}-2\alpha_{.23})]]$

(A.44) $\mathcal{E}[C^*(Y_1,Y_2,V)]^2 = \alpha_{.22}/8+3/2$

(A.45) $\mathcal{E}[F^*(Y_1,Y_2,V)]^2 = [(\alpha_{22}-\alpha_{33})^2+8(\alpha_{22}\alpha_{33}-\alpha^2_{23})$

$$+48(\alpha_{22}+\alpha_{33}-\alpha_{23})]/8(\alpha_{22}+\alpha_{33}-2\alpha_{23})$$

$$-3(\alpha_{22}-\alpha_{33})^2/2(\alpha_{22}+\alpha_{33}-2\alpha_{23})^2$$

(A.46) $\mathcal{E}[H^*(Y_1,Y_2,Y_3,V)]^2 = [-(\alpha_{22}-\alpha_{23})^2+6(\alpha_{33}+\alpha_{23})-6\alpha^2_{23}/$

$$\alpha_{22}-6(\alpha_{22}-\alpha_{23})^2/(\alpha_{22}+\alpha_{33}-2\alpha_{23})$$

$$+(\alpha_{22}-\alpha_{23})^2(\alpha_{22}-\alpha_{23}+6)/(\alpha_{22}\times$$

$$(\alpha_{22}+\alpha_{33}-2\alpha_{23}))]/\alpha_{22}(\alpha_{22}+\alpha_{33}$$

$$-2\alpha_{23})$$

(A.47) $\mathcal{E}C^*(Y_1,Y_2,V)F^*(Y_1,Y_2,Y_3,V)$

$$= -[2(\alpha^2_{22}-\alpha^2_{23})+12(2\alpha_{22}-\alpha_{23})+(\alpha_{22}-\alpha_{33})(\alpha_{22}-\alpha_{23})$$

$$(\alpha_{22}-\alpha_{23}+6)/(\alpha_{22}+\alpha_{33}-2\alpha_{23})]/[8\alpha^{\frac{1}{2}}_{22}(\alpha_{22}+\alpha_{33}$$

$$-2\alpha_{23})^{\frac{1}{2}}]$$

(A.48) $\mathcal{E}C^*(Y_1,Y_2,V)H^*(Y_1,Y_2,Y_3,V)$

$$= (\alpha_{22}-\alpha_{23}+6)[1-(\alpha_{22}-\alpha_{23})^2/\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})]/$$

$[4(\alpha_{22}+\alpha_{33}-2\alpha_{23})^{\frac{1}{2}}]$

(A.49) $\mathcal{E}F^*(Y_1,Y_2,Y_3,V)H^*(Y_1,Y_2,Y_3,V)$

$$= [-\alpha^2_{22}+\alpha_{22}\alpha_{23}+3\alpha_{22}\alpha_{33}+\alpha_{33}\alpha_{23}-4\alpha^2_{23}+6(\alpha_{22}+3\alpha_{33}) +2\alpha_{23}(\alpha_{22}-\alpha_{23})(\alpha_{23}+6)/\alpha_{22}+12(\alpha_{22}-\alpha_{33})(\alpha_{22}-\alpha_{23})/ (\alpha_{22}+\alpha_{33}-2\alpha_{23})+(\alpha_{22}-\alpha_{33})(\alpha_{22}-\alpha_{23})^2(\alpha_{22}-\alpha_{23}+6)/ (\alpha_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23}))]/[4\alpha^{\frac{1}{2}}_{22}(\alpha_{22}+\alpha_{33}-2\alpha_{23})].$$

REFERENCES

[1] Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley, New York.

[2] Anderson, T.W. (1966). Some nonparametric multivariate procedures based on statistically equivalent blocks. Proc. 1st Internat. Symp. Mult. Analysis, 5-27.

[3] Anderson, T.W. (1973). An asymptotic expansion of the distribution of the studentized classification statistic W. Ann. Statist. 1, 964-972.

[4] Berk, R.H. (1966). Limiting behavior of posterior distributions when the model is incorrect. Ann. Math. Statist. 37, 51-58.

[5] Bickel, P.J. and Yahav, J.A. (1968). Asymptotically optimal Bayes and minimax procedures in sequential estimation. Ann. Math. Statist. 39, 442-456.

[6] Chanda, K.C. and Lee, J.C. (1975). A class of nonparametric classification rules. To appear in Some Statistical Methods Useful in Oil Exploration. Ed. D. B. Owen.

[7] Chow, Y.S. and Robbins, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. Ann. Math. Statist. 36, 457-462.

[8] Cover, T.M. (1968). Estimation using the nearest neighbor rule. IEEE Trans. Inform. Theory, IT-14, 50-57.

[9] Cover, T.M. and Hart, P.E. (1967). Nearest neighbor pattern classification. IEEE Trans. Inform. Theory, IT-13, 21-26.

[10] Das Gupta, S. (1964). Nonparametric classification rules. Sankhya A, 26, 25-30.

[11] Das Gupta, S. and Kinderman, A. (1974). Classifiability and designs for sampling. Sankhya A, 36, 237-250.

[12] Fix, E. and Hodges, J.L. (1951). Nonparametric discrimination: Consistency Properties. U. S. Air Force School of Aviation Medicine, Report No. 4, Randolph Field, Texas.

[13] Fraser, D.A.S. (1957). Nonparametric Methods in Statistics. John Wiley, New York.

[14] Goldstein, Matthew (1962). K_n-nearest neighbor classification. IEEE Trans. Inform. Theory, IT-18, 627-630.

[15] Grams, W.F. and Serfling, R.J. (1973). Convergence rates for U-statistics and related statistics. Ann. Statist., 1, 153-160.

[16] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Ann. Math. Statist., 19, 293-325.

[17] Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Institute of Statistics Mimeo Series No. 302, University of North Carolina, Chapel Hill, N.C.

[18] Hoeffding, W. (1963). Probability inequality for the sums of bounded random variables. Jour. Amer. Statist. Assoc., 58, 13-30.

[19] Hoeffding, W. and Wolfowitz, J. (1958). Distinguishability of sets of distributions. Ann. Math. Statist., 29, 700-718.

[20] Hudimoto, H. (1964). On a distribution-free two-way classification. Ann. Inst. Statist. Math., 9, 31-36.

[21] Lehmann, E.L. (1951). Consistency and unbiasedness of certain nonparametric tests. Ann. Math. Statist., 22, 165-179.

[22] Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function. Ann. Math. Statist., 36, 1049-1051.

[23] Sen, P.K. (1960). On some convergence properties of U-statistics. Bull. Calcutta Statist. Assoc., 10, 1-18.

[24] Simons, G. (1968). On the cost of not knowing the variance when making a fixed-width confidence interval for the mean. Ann. Math. Statist., 39, 1946-1952.

[25] Srivastava, M.S. (1973). A sequential approach to classification: Cost of now knowing the covariance matrix. Ann. Mult. Analysis, 3, 173-183.

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report No. 276	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) On Some Aspects of the Classification Problem		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Hsien Elsa Lin		8. CONTRACT OR GRANT NUMBER(s) DAAG-29-76-G-0038
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Theoretical Statistics University of Minnesota ~~Minneapolis, Minnesota~~ 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE October 1976
		13. NUMBER OF PAGES 116
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE NA

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Classification, Multivariate Normal, probability of misclassification, asymptotic expansion, nearest neighbor rules, sequential rules, U-statistics.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Different rules for the problem of classifying one (or more) unit to one of several distinct populations on the basis of random training samples are considered and their asymptotic probabilities of misclassification (PMC) are derived.

First, we consider the problem of classifying one unit to one of three distinct multivariate normal distributions with a common covariance matrix. A plug-in rule is obtained by replacing the unknown parameters by their estimates in the optimal classification rule. Following

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20., cont...

T. W. Anderson, we obtain the asymptotic expansions of the PMC's and the estimated PMC's of this plug-in rule.

In the second chapter, we consider the problem of classifying one unit to one of two distinct populations with completely unknown distribution functions. The usual nearest neighbor (NN) rules can't be applied if the observations are available only in their relative orders or ranks. Using the basic ideas of NN rules, we propose some rules expressed in terms of their ranks and derive the asymptotic PMC's of the rules.

In the third chapter, rules based on U-statistics are suggested and the asymptotic PMC of the rules are obtained together with the rate of convergence as the sizes of the training samples approach infinity.

Finally, we consider sequential rules based on U-statistics in order to control the PMC uniformly and arbitrarily. The moment generating function is shown to exist. The proof of asymptotic properties of the sequential rules suggested by Srivastava for the classification into one of two multivariate normal distributions is rigorized and corrected.

[26] Widder, D.V. (1961). Advanced Calculus. 2nd Edition, Prentice-Hall, Inc., N. J.